Untangling the West-Coastal Bantu Mess : Identification, Geography and Phylogeny of the Bantu B50-80 Languages

Pacchiarotti, Sara ; Chousou-Polydouri, Natalia ; Bostoen, Koen

Abstract: In this paper we deal with the identification, geography and internal phylogeny of Bantu B50-80 languages. According to lexicostatistical and lexicon-based phylogenetic studies, these languages belong to the West-Coastal branch of the Bantu family along with the Kikongo Language Cluster. First, we present an updated list of nearly 100 language varieties falling into Guthrie's B50-80 groups along with corresponding updated geographical coordinate values. We dissipate confusion surrounding some glossonyms and point out misinterpretations that led to conflation of different varieties and misnumbering in non-genetic, referential classifications. Second, we present the results of a new phylogenetic study including all B50-80 varieties in our sample to be later compared to an internal classification based on shared phonological innovations. Our results show that: (i) previous internal subgroupings of Guthrie's B50-80 languages within West-Coastal Bantu either need to be revised or are no longer valid against new evidence; and (ii) the new internal structure of the West-Coastal branch suggests that the homeland of Proto-West-Coastal Bantu speakers is not to be found, as previously believed, somewhere in between the Bateke Plateau and the Bandundu region in the DRC but rather much more eastward, i.e. somewhere between the Kamtsha and Kasai Rivers in the DRC. = Dans cet article, nous traitons de l'identification, de la géographie et de la phylogénie interne des langues bantu B50-80. Selon les études lexicostatistiques et phylogénétiques basées sur le lexique, ces langues appartiennent à la branche Côte-Ouest de la famille bantu, tout comme le groupe kikongo. Nous commençons par présenter une liste mise à jour des près de 100 variantes linguistiques des groupes B50-80 de Guthrie et des coordonnées géographiques correspondantes, mises à jour également. Nous dissipons la confusion qui entoure certains glossonymes et pointons du doigt les mauvaises interprétations qui ont conduit au regroupement de variétés en réalité distinctes, et à des erreurs de numérotation dans les classifications référentielles non génétiques. Nous présentons ensuite les résultats d'une nouvelle étude phylogénétique de toutes les variantes B50-80 de notre échantillon, que nous comparerons plus tard à une classification interne réalisée sur base d'innovations phonologiques partagées. Les résultats que nous avons obtenus montrent que (i) les classifications internes des langues B50-80 doivent être révisées, soit être reconsidérées entièrement et (ii) la structure interne de la branche Côte-Ouest se dégageant de notre étude suggère que son berceau ne se trouve pas, comme on l'a cru précédemment, quelque part entre le plateau Batéké et la région du Bandundu en RDC, mais beaucoup plus à l'est, entre les rivières Kamtsha et Kasaï, toujours en RDC.

DOI: https://doi.org/10.2143/AL.25.0.3287234

Originally published at:
Pacchiarotti, Sara; Chousou-Polydouri, Natalia; Bostoen, Koen (2019). Untangling the West-Coastal Bantu Mess : Identification, Geography and Phylogeny of the Bantu B50-80 Languages. Africana Linguistica, 25:155-229.
DOI: https://doi.org/10.2143/AL.25.0.3287234

Untangling the West-Coastal Bantu mess: identification, geography and phylogeny of the Bantu B50-80 languages

Sara Pacchiarotti, Natalia Chousou-Polydouri and Koen Bostoen

Abstract

In this paper we deal with the identification, geography and internal phylogeny of Bantu B50-80 languages. According to lexicostatistical and lexicon-based phylogenetic studies, these languages belong to the West-Coastal branch of the Bantu family along with the Kikongo Language Cluster. First, we present an updated list of nearly 100 language varieties falling into Guthrie's B50-80 groups along with corresponding updated geographical coordinate values. We dissipate confusion surrounding some glossonyms and point out misinterpretations that led to conflation of different varieties and misnumbering in non-genetic, referential classifications (Guthrie 1971; Maho 2009; Hammarström 2019). Second, we present the results of a new phylogenetic study including all B50-80 varieties in our sample to be later compared to an internal classification based on shared phonological innovations. Our results show that: (i) previous internal subgroupings of Guthrie's B50-80 languages within West-Coastal Bantu either need to be revised or are no longer valid against new evidence; and (ii) the new internal structure of the West-Coastal branch suggests that the homeland of Proto-West-Coastal Bantu speakers is not to be found, as previously believed, somewhere in between the Bateke Plateau and the Bandundu region in the DRC but rather much more eastward, i.e. somewhere between the Kamtsha and Kasai Rivers in the DRC.

Keywords: West-Coastal Bantu, West-Western Bantu, Bantu B50-60-70-80, Kikongo Language Cluster, glossonyms, Guthrie's referential classification, linguistic geography, phylogenetics, internal classification, subgrouping, Bantu Expansion, homeland.

Introduction

West-Coastal or West-Western Bantu languages are spoken in West-Central Africa, more specifically in parts of Gabon, the Republic of the Congo, the Democratic Republic of the Congo (henceforth DRC) and northern Angola. They constitute a discrete branch within the Bantu family. We consider them to be the descendants of the language(s) spoken by the first Bantu speech communities that settled south of the Equatorial rainforest about 2500 years ago after an inland expansion through the central forest block (cf. Bostoen *et al.* 2015; Grollemund *et al.* 2015). The material culture, subsistence and language dynamics of these early Bantu-speaking settlements are the main focus of the on-going BantuFirst project.[1] This article is meant as a state-of-the-art reference for historical-comparative linguistic research within this cross-disciplinary project and hopefully for future research on West-Coastal Bantu (henceforth WCB) languages.

The genealogical unity of the WCB branch – or West-Western as Grollemund *et al.* (2015) call it – has been established through lexicostatistical (Vansina 1995; Bastin *et al.* 1999) and lexicon-based phylogenetic studies (de Schryver *et al.* 2015; Grollemund *et al.* 2015; Bostoen & de Schryver 2018a, b). In these phylogenetic studies, the WCB clade splits into three major subgroups: (1) the Kikongo Language Cluster (KLC), comprising Guthrie's B40, H10 and H30 groups plus Hungan (H42) and Samba (L12a); (2) the Nzebi-Mbete-Teke subgroup consisting predominantly of Guthrie's B50, B60 and B70 groups and (3) the Yanzi subgroup, including most of Guthrie's B80 group. However, the sample of WCB languages in previous research is limited, especially for subgroups (2) and (3), because earlier studies targeted either the entire Bantu family (Grollemund *et al.* 2015) or a specific subgroup (i.e. the KLC) (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a, b). One of the main objectives of this article is to propose an internal lexicon-based phylogenetic classification of WCB incorporating as many Guthrie's B50-80 languages as possible in order to assess whether previous subgrouping holds against further scrutiny.

However, in order to achieve this goal, we first need a clear picture of which varieties could be part of WCB outside of the KLC. Existing referential classifications of Bantu languages such as Guthrie (1948, 1953, 1970, 1971), Maho (2003, 2009) and Hammarström (2019) do not include all varieties potentially belonging to Bantu B50-80 groups, erroneously lump varieties which are sometimes referred to by the same glossonym, or assign the same Guthrie's alphanumeric code to two distinct varieties. Second, we need to have an understanding of the geographical distribution of Bantu B50-80 languages in order to determine the distribution of their linguistic features and uncover the historical changes they underwent. The lexicon-based phylogenetic classification proposed in this article can then be compared against shared phonological, morphological and syntactic innovations identified in future comparative research.

1. BantuFirst is a 5-year program (2018-2022) funded by a Consolidator's Grant (n° 724275) of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, which was granted to the last author. For more details, see bantufirst.ugent.be.

In line with these three objectives, i.e. the identification, geography and phylogeny of Bantu B50-80 languages, this article is structured as follows. In Section 1, we discuss issues concerning the identification and geography of language varieties falling into Guthrie's B50-80 groups. We single out varieties that have not been inventoried in the referential classifications of Guthrie (1948, 1953, 1970, 1971) or its updates by Maho (2003, 2009) or that have been mistakenly lumped together under the same alphanumeric code but should be kept separate. In Section 1.1, we explain how we assigned glossonyms to the varieties included in this study. In Section 1.2, we explain the rationale behind the alphanumeric codes we assigned to varieties in this study. In Section 1.3, we offer updated geographical coordinate values for each variety and plot them on a map. In Section 2, we untangle confusion observed in the literature around the following glossonyms: Mbete and/or Mbaama (Section 2.1), Boma and/or Buma (Section 2.2), Mpuono, Mpuun and/or Mbuun (Section 2.3), Teke, Tyo and/or Tio (Section 2.4) and Mfununga, Mfunika, Mfunuka, etc. (Section 2.5). In Section 3, we present the results of a lexicon-based phylogenetic study of B50-80 varieties and show that previous phylogenetic subgroupings (i.e. Nzebi-Mbete-Teke B50-70 vs. Yanzi B80) do not hold against new evidence. Our latest understanding of WCB internal genealogical structure also leads to a new hypothesis about the location of its homeland. While it was situated earlier "somewhere in between the Bateke Plateau and the Bandundu region" (Bostoen *et al.* 2015: 361), we propose here that the homeland is rather in the area between the Kamtsha and the Kasai Rivers southeast of Bandundu (city) in the current Kwilu province of the DRC. In the Appendix, we offer an updated list of nearly 100 varieties falling into Guthrie's B50-80 groups including an in-depth discussion of topics dealt with in Section 1 (see A.1, A.2 and A.3).

Throughout this paper we use the term "(language) variety" as "a neutral term to apply to any particular kind of language which we wish, for some purpose, to consider as a single entity" (Chambers & Trudgill 1998: 5). All varieties in our dataset are "doculects". By doculect we mean a language variety that has been documented in a particular source, regardless of the amount of extant documentation on that particular variety (Bowern 2008: 8; Cysouw & Good 2013). Therefore, we use the term "(language) variety" as a synonym of "doculect". Different doculects can represent one and the same "dialect", which we use here not pejoratively but rather as a synonym of "regiolect", i.e. a geographically determined way of speaking.

1. Identification and geography of B50-80 varieties

Our purpose in this section is not to offer an overview of the extant literature on Guthrie's B50-80 varieties. Nevertheless, we briefly note that extensive lists of published and unpublished linguistic, historical, ethnographic and anthropological works on Guthrie's B70-80 varieties spoken in the DRC can be found in van Bulck (1948), Boone (1973) and Maalu-Bungi *et al.* (2011). Reference materials for B50-70 languages spoken in Gabon include Mouguiama-Daouda (2005), Idiata (2007), and Idiata *et al.* (2013). Reference materials for B60-B70 languages spoken in the Republic of the Congo include Lane (1989) and Leblanc *et al.* (2012).

Variety	Secondary sources
Wanzi (B501)	Hombert and Mouele (1988)
Duma (B51)	Mickala-Manfoubi (1988)
Nzebi (Gabon) (B52)	Marchal-Nasse (1989)
Tsaangi (B53)	Loubelo (1987)
B50 in general	Mouele (1997)
Nduumo (B63)	Biton (1969), Lane (1989)
Mbaama (Gabon) (B62)	Okoudowa (2016)
Mbete (Congo) (B61)	Lane (1989)
Ngungwel (B72a)	Rurangwa (1982), Paulian (1994)
Laali (B73b)	Bissila (1991)
Yaa (B73c)	Mouandza (2001)
Eboo-Nzikou (B74)	Raharimanantsoa (2012a, b, 2017)
Fumu (Congo) (B77bX)	Makouta-Mboukou (n.d., 1976)[2]
Boma Yumu (B80zX)	Hochegger (1972a), Burssens (1999)
West Ding (B86U)	Mertens (1939), Ebalantshim Masuwan (1980)
East Lwel (B862X)	Khang Levy (1979)
Tiene (B81X)	Ellington (1977)
East Yans (B85b)	Impubi Mukwa (1987)
Mpur (B85eY)	Kibwenge India'Ane (1985)
Nsambaan (B85FX)	Mfum-Ekong (1979)
Nzadi (B865X)	Crane *et al*. (2011)
Mbuun (B87)	Mundeke (1977, 1979), Dibata Mimpiya (1977)
Nsong (B85dZ), Ngong (B864X), Mpiin (B863Y), Mbuun Imbongo (B87W), East Ngwi (B861X)	Koni Muluwa (2010)

Table 1. Secondary sources for some B50-80 varieties in this study

The lexical data on language varieties belonging to Guthrie's B50-80 groups included in this study are second-hand and come from five main sources: (i) the B50-80 lists of "basic vocabulary" used for the seminal lexicostatistical study of Bastin *et al.* (1999); (ii) the comparative lexicon of Bantu languages spoken in the Kwilu region of the DRC (mostly Guthrie's B80) in Koni Muluwa & Bostoen (2015); (iii) the wordlists in Nsuka Nkutsi (1990) on Boma Nkuu (B80x), South Boma [Nkuu] (B80y), Ngi (B70y) and Bibaana (B70x); (iv) the grammatical description of Stappers (1986) for the North Boma (B82) variety spoken in Mushie, in the Mai Ndombe province of the DRC; and (v) a wordlist of Vili (B503) from the Laboratoire Dynamique du Langage (DDL) at the University of Lyon 2.

2. According to OLAC Language Resource Catalog, Makouta-Mboukou (n.d.) is dated (1960), while the Glottolog dates it (1969). We do not know which date is right because the typewritten manuscript itself, kindly provided to us by Barbara Thomas from Dallas International University, has no date.

Additionally, we used the secondary sources in Table 1 to fill in gaps in the data and/or refine our understanding of the geo-linguistic situation of a particular variety.[3]

We tried to include as many works from African universities as possible. Nevertheless, we know that there are many more references on B50-80 in their countries of origin which we were unfortunately unable to access.

Table 2 in the Appendix offers a complete list of doculects used in this study. In the Appendix, we explain in detail how Table 2 is organized. In the following sections we discuss the contents of several columns in Table 2 that are immediately relevant to the discussion at hand. Sections 1.1 and 1.2 deal with variety names and alphanumeric codes respectively. In Section 2, we attempt to dissipate dangerous confusion on several glossonyms and their alternative names.

1.1. Rationale behind variety names assigned in this study

By comparing the second and third columns in Table 2 in the Appendix,, the reader will notice that in the vast majority of cases we modified variety names with respect to the original source. Whenever we did so, we attempted to be maximally informative and minimally confusing. We strived to achieve these two goals in the following ways: (i) by including in the variety name an identifiable and retrievable location (village, town, city, region, country) where the variety is spoken, e.g. Nduumo (Yéyé) (see A.1 and Section 1.3 for further discussion); (ii) by adding regiolectal information on a doculect whenever available, e.g. East Lwel (Sedzo) (see A.2); and (iii) by dissipating the confusion surrounding some glossonyms (see Section 2).

Our choice of a glossonym instead of another for any of the varieties in Table 2 is unfortunately rather arbitrary. In general, we preferred endonyms to exonyms (cf. for instance *Mbuun* instead of *Mbunda*, *Yans* instead of *Yanzi*, *Ding* instead of *Dinga*) but we did not apply this preference consistently across the board (e.g. *Wuumu* B78 is an exonym, *Wuũ* is the endonym, but *Wuumu* appears to be considerably more frequent than *Wuũ* to refer to B78 in the literature).[4] Alternative names for varieties listed in Table 2 can be found in Guthrie (1970, 1971), Maho (2009), Idiata *et al.* (2013), Maalu-Bungi *et al.* (2011) and the Ethnologue (Lewis *et al.* 2016), among others. Invaluable sources for alternative names of varieties spoken in the DRC are Maes & Boone (1935), van Bulck (1948, 1954) and Boone (1973).

3. In this paper, we follow Maho (2009) in writing names of Bantu languages without a noun class prefix, e.g. Yaa (B73c) instead of Iyaa. Alphanumeric codes deviating from the standard Guthrie format of a letter plus a number, e.g. Boma Yumu (B80xZ), will be explained in Section 1.2.

4. At times, the available information is partially conflicting. For example, according to Boone (1973: 349), the Bayansi found in the mission of Mbeno call themselves Bayey (but also Bayay, Bayey, Bayoy) in their own language. Others (not from Mbeno) want to be called only with the Kongo exonym "Bayansi". On the other hand, Swartenbroeckx (1948: 2) says that Bayey or Bayay is simply a contraction of the more southern term Yenzi or Yansi.

1.2. Rationale behind alphanumeric codes assigned in this study

In Table 2 in the Appendix there are two columns for alphanumeric codes, one which follows the conventions in Guthrie (1971) and updates in Maho (2009), and one with codes we assigned to the varieties in this study. As with variety names, we attempted to be maximally informative and minimally confusing in assigning "modified" alphanumeric codes. The conventions we used for this purpose are as follows:

a. Decimal number where the second digit is different from zero followed by (uppercase) X, Y, Z, etc., e.g. North Boma (Mushie) (B82X), North Boma (Mbali-Iboma) (B82Y). With capital X, Y, Z, we indicate that we have data on doculects inventoried in Guthrie (1971) and/or Maho (2009) from more than one geographical location and that we consider these to be regiolectal varieties of the same language variety, i.e. B82X and B82Y are regiolects of B82.[5] We used this convention also with data from one single location whenever we were able to gather additional dialectal information on that doculect, cf. #78 East Lwel (Sedzo) (B862X), where X indicates the eastern variety of Lwel.
b. Decimal number where the second digit is zero followed by (lowercase) x, y, z, etc., e.g. Boma Nkuu (Monkana) (B80x), South Boma [Nkuu] (Boku) (B80y). We use this convention to indicate that a variety is not inventoried in either Guthrie (1971) or Maho (2009) and we tentatively place it in one of Guthrie's groups. For example, lowercase x in B80x indicates that Boma Nkuu is a variety not previously inventoried; B80 in B80x means that we are tentatively placing it within Guthrie's B80 group based on its geographical location. Lowercase x and y in B80x and B80y also mean that at the present time we consider Boma Nkuu (Monkana) (B80x) and South Boma [Nkuu] (Boku) (B80y) to be two distinct languages, rather than dialectal variants of a single language.

The conventions in a) and b) can also be combined, cp. #44 Boma Yumu (Pentane/ Mondai) (B80zX) and #45 Boma Yumu (Saio) (B80zY). B80z means that both varieties are instances of the uninventoried language Boma Yumu. Capital X and Y after B80z mean that #44 and #45 are two distinct regiolects of Boma Yumu. Conventions in a) and b) can also be combined with Guthrie (1971)/ Maho (2009) codes, e.g. #71 Mpur (Due) (B85eY) and #72 Mpur (Kolonzadi) (B85eZ). Capitals Y and Z after the Maho (2009) code for Mpur (B85e) mean that (B85eY) and B85eZ are two regiolects of B85e.

1.3. An update of geographical coordinates for B50-80 varieties.

In this section we deal with the geographical location of B50-80 varieties presented in Table 2. Their geographical distribution is shown in Map 1. Map 1 is essential to the study of B50-80 groups as their geography can be informative to historically

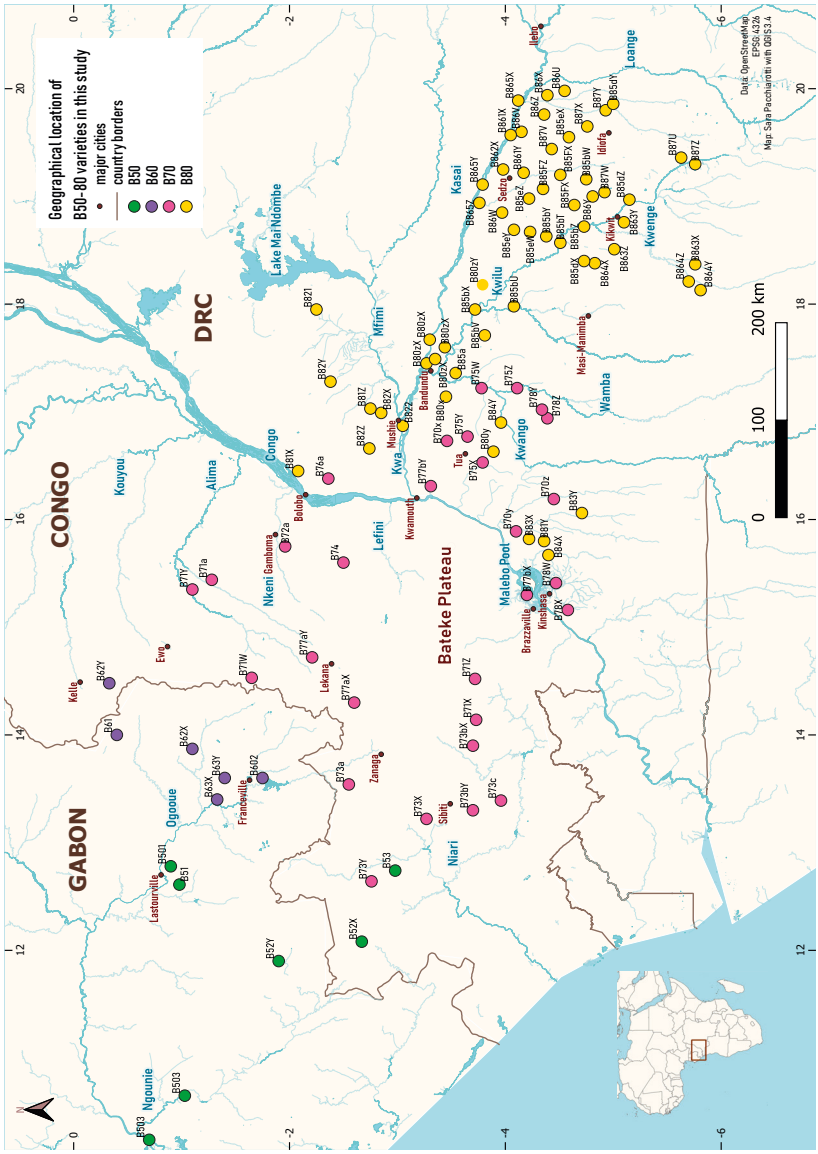
5. We opted for uppercase letters towards the end of the alphabet (e.g. X, Y, Z or further back if we have more than three distinct varieties, e.g. W, V, U, etc.) to avoid creating confusion and/or overlap with Maho (2009), where uppercase letters at the beginning of the alphabet (e.g. A, B, C) are used to distinguish different dialectal varieties of a language.

interpret the distribution patterns of certain linguistic features across WCB. Depending on their geographical spread, phonological or other linguistic changes affecting one or more groups can be analyzed as the most probable result of either inheritance, language contact or parallel innovation, or still as a combination of these scenarios. First, we discuss how we updated the geographical coordinate values in Table 2 and how we dealt with problems and mismatches between place names and coordinate values in Bastin *et al.* (1999) (see A.1 for a detailed discussion). Then we plot these values on Map 1.[6]

When we found no coordinate values for a doculect in the original source, we obtained them from geonames.org, a geographical database with over 11 million placenames, by searching place names as indicated in the original source. If a source specified more than one place name and we were able to find these on geonames.org, we included more than one set of coordinates for a given doculect (see for instance #44 and #73).

When we found coordinate values in Bastin *et al.* (1999), we systematically cross-checked them against the corresponding place names by using geonames.org and the website gps-coordinates.net. In general, we provided slightly more precise coordinates with two decimals after the point. For each entry we made sure that the place name of a doculect could be identified and retrieved from geonames.org (see A.1). If this were not the case, we started our cross-checking from the coordinate values found in Bastin *et al.* (1999) by using the website gps-coordinates.net and then searched for alternative place names retrievable on geonames.org with the closest possible matching set of coordinate values. When Bastin *et al.* (1999) provide only the name of a country and a set of coordinates, we used gps-coordinates.net to roughly locate the dot corresponding to the coordinate set and then refined the search for a place close to that coordinate set on geonames.org. While doing this inspection, we spotted several problems and mismatches between place names and coordinates in Bastin *et al.* (1999). We address these in detail in A.3 in the Appendix.

6. An interactive version of Map 1 can be found on the BantuFirst project website: bantufirst.ugent.be/research/west-coastal-bantu-interactive-map.



Map 1. Indicative geographical distribution of B50-80 varieties surveyed in this study

2. Untangling confusion around language variety names

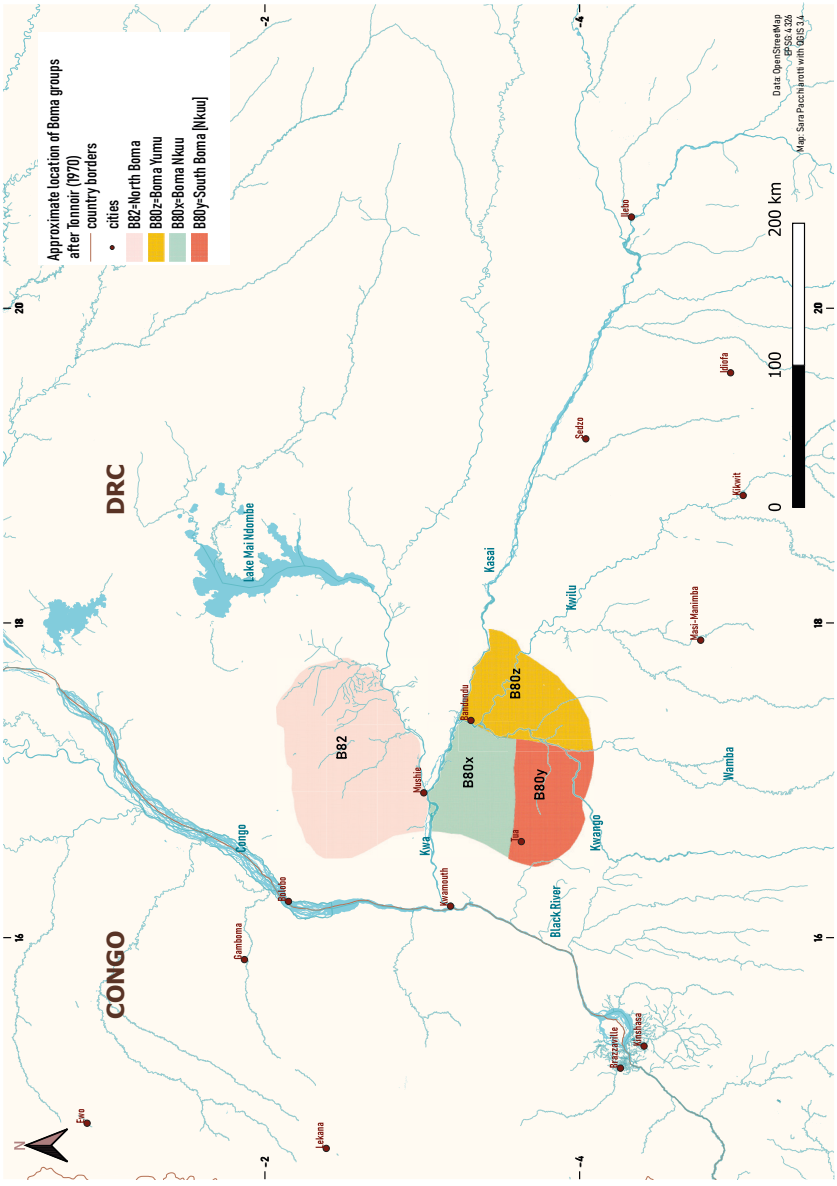
In this section we discuss some glossonyms in Guthrie's B60-80 groups which are prone to confusion and/or in urgent need of clarification. These are: Mbete and/or Mbaama (see #8-10 in Table 2), Boma and/or Buma (see #42-45 and #51-53 in Table 2), Mpuono, Mpuun and/or Mbuun (see #56-57 and #94-99 in Table 2), Teke, Tyo and/or Tio (see #14 and #16-41 in Table 2), and Mfununga, Mfunika, Mfunuka etc., alternative names used to refer to several distinct varieties spoken mainly in the DRC (such as #36-41 and #54-55 in Table 2). We discuss each of these in turn in the following subsections.

2.1. Mbete and/or Mbaama

Mbete (aka Mbede, Mbere) can refer to a group of Bantu languages spoken in Gabon and the Republic of the Congo, i.e. Guthrie's B60 group (cf. Raponda-Walker 1960: 9; Jacquot 1978: 495), or to a specific language within this group, namely B61. Maho (2009) lists the following languages and alternative names under B60 Mbete group: B602 Kaning'i, B603 Yangho (Yongho), B61 Mbete (Mbere), B62 Mbaama (Mbamba) including Mpiini B601, and B63 Nduumo (Mindumbu). A map showing the exact geographical location of B60 languages in Gabon and the Republic of the Congo can be found in Lane (1989: 11). Lane (1989: 7) lists the following as alternative names for Mbete B61: Mbeti, Mbede, Mbere, Limbede, Obamba, Mbama, Mbamba, Lembaamba, Mbaama, Bambaama, and Gimbaama. According to Lane (1989), in the southern part of the Republic of the Congo, Mbete (B61) is referred to as Mbamba. Mbamba is also one of the glossonyms of B62 (cf. supra). Lane (1989) also reports that according to some sources, Mbete people (presumably people speaking one of the B60 languages) are officially known in Gabon as Obamba, a French rendering of Mbamba. This information is confirmed by the nineteenth edition of the Ethnologue (Lewis *et al.* 2016), where Mbaama is given as alternative name for both Mbete (B61) and Ombamba (B62).

2.2. Boma and/or Buma

Tonnoir (1970: 2 and ff.) identifies four distinct groups in the DRC of what he calls in French *(Ba-)boma* people. Their approximate geographical location is shown in Map 2. These groups are often considered as sub-entities of one and the same people (Roelants 1969: 24). In some maps, these different groups are all named *Boma* (Soret 1955).



Map 2. Approximate location of distinct Boma groups (based on Tonnoir 1970: 2)

The first group described by Tonnoir (1970) is called *(Ba-)boma nord* and located north of the Kwa River in the DRC east of Bolobo, around Mushie and towards Lake Mai-Ndombe, in the current DRC province of Mai-Ndombe (see B82 in Map 2). This same group and/or their language is alternatively labeled *Giribuma* (Vossius 1666),[7] *(Ba-)boma Badia* (Verhulpen 1936) (probably because they have as eastern neighbors the Badia, a.k.a Badja, Bajia C34 speakers), *Wabuma* (Johnston 1884; Maes 1924; Maes and Boone 1935: 46), *Boma-Nuni* (Roelants 1969: 24) or *(ki-~ke-)Boma/Buma* (Johnston 1908, 1922; Focquet & van der Kerken 1924; van Bulck 1948, 1954; Detienne 1956; Storme 1956; Vansina 1966 among others). The language spoken by this group has been documented by Stappers (1986) who calls it in German *Boma* but includes in his work (Stappers 1986: VI) a map from Sulzmann (1983: 532) where, following Tonnoir (1970), the area where the language is spoken is labeled *Baboma nord* (see also Van Acht 1949: 20). Stappers (1986) states that the Guthrie code assigned to the language he describes is B82. This is also what we find in Guthrie (1953: 81), who says that the language he calls Boma (B82) is "spoken in Congo Belge […] inland from Mushie on the north bank of the R[iver] Fimi and the R[iver] Kwa." (see Map 2 and map in Guthrie 1953).[8] We renamed Boma (B82) in Guthrie (1953) North Boma (B82) after Tonnoir (1970) and Sulzmann (1983) to minimize confusion with other varieties which are also called Boma. The Glottolog (Hammarström *et al.* 2018) assigns to North Boma (B82) the glottocode [boma 1246] and the ISO 639-3 code [boh]. The Glottolog erroneously reports that this variety is spoken in the Republic of the Congo and the DRC (Hammarström *et al.* 2018). The geographical coordinates found in the Glottolog for North Boma (B82) are also wrong: they point to a place in Cameroon, while North Boma (B82) is spoken only in the DRC.

The second group identified by Tonnoir (1970: 4) is called *(Ba-)boma Yumu* (cf. also Focquet & van der Kerken 1924) and is located south of the Kasai River in the DRC, southwards and southeastwards of Bandundu (city) (formerly Banningville) in the current DRC province of the Kwilu (see B80z in Map 2). Vansina (1973-1974: 338) calls this group *Boma Boka*, while Selvaggi (1968) refers to it as *Boma de Bandundu*. In the Ethnologue (Grimes 1992: 89; Lewis *et al.* 2016: 127, 388), what we call Boma Yumu after Tonnoir (1970) appears as Boma with alternative names Buma, Kiboma, Boma Kasai, cf. also Johnston (1908: 564), who uses *Baboma* or *Babuma of the Kwa-Kasai*. The language of this group comprises at least four distinct dialectal varieties (Hochegger 1972a: III). Strictly linguistic documentation on Boma Yumu includes two dictionaries (Hochegger 1972a; Burssens 1999) and

7. Tonnoir (1970) explains that the exonym *Giribuma* was assigned to this group of people by Dapper, a 17th c. Dutch geographer and doctor. *Giribuma* is probably the phonetic deformation of *Ngeliboma*, the name given to a sovereign of the Baboma during early European explorations.

8. In Guthrie's (1948) grouping of zone B, there is a language called BOMA (uppercase in the original) and numbered 42 (cf. Table 3). This language is grouped under group 40 together with Mfinu (B83), Tiene (B81) and Yans (B85) (among others) and is likely to be equivalent to Boma B82 in Guthrie (1953). There is also a language called Boma (lowercase in the original), numbered 33 (Guthrie 1948: 33) and grouped under group 30 together with Fumu (B77), Teke (B71) and Ngee (B76). This lowercase Boma very likely refers to current Eboo-Nzikou B74, a Teke language spoken in the Republic of the Congo.

collections of oral tradition texts (Hochegger 1972b, 1976). Both Hochegger (1972a, b; 1976) and Burssens (1994, 1999) call this language *buma*. Burssens (1994: 72, 1999: i) erroneously assigns to it the Guthrie code B82, that is, the same code used for North Boma (cf. supra). In his inventory of Bantu languages, Hammarström (2019: 26) too assigns the code B82 to what we call Boma Yumu and what he calls Boma (Kiboma, Boma Kasai, Buma), perhaps subsuming North Boma and Boma Yumu under one single code. By comparing the synchronic sound system, the diachronic sound changes and the vocabulary of the language described by Stappers (1986), which we call North Boma, with those of the language described by Hochegger (1972a) and Burssens (1999), which we call Boma Yumu, it is clear that these two doculects cannot be considered the same language, and not even regiolectal varieties of a single language. Because Boma Yumu is not inventoried in Guthrie (1948, 1953, 1971) or Maho (2003, 2009), we assigned to it the code B80z. B80 in B80z means that this language belongs to Guthrie's B80 group geography-wise; z in B80z means that Boma Yumu is a distinct language within B80 and should be assigned its very own alphanumeric code within this group (cf. Section 1.2 for a detailed discussion of our Guthrie-inspired alphanumeric codes). There is no glottocode or ISO 639-3 code for Boma Yumu in the Glottolog (Hammarström *et al.* 2018), but the Ethnologue (Eberhard *et al.* 2019) gives Boma Yumu (B80z) the ISO 639-3 code [boh], i.e. the same ISO code used by the Glottolog (Hammarström *et al.* 2018) for North Boma (B82).

The third group distinguished by Tonnoir (1970: 4) is called *(Ba-)boma (Ba-)nkuu* and is located in the DRC south of the Kwa River and west of the Kwango River in the vast area northeast of Kinshasa (see B80x in Map 2). This group is also reported in Sulzmann (1983: 532) and Stappers (1986: 131), but the extent of its geographic distribution varies from source to source. To our knowledge, the only existing documentation on the language spoken by the Boma Nkuu is a wordlist found in Nsuka Nkutsi (1990) who calls this language *nkuu*. Based on the extremely scanty information available to us, we believe Boma Nkuu to be yet a different language with respect to North Boma (B82) and Boma Yumu (B80z) (cf. also Hochegger 1972a: III for a similar observation). Boma Nkuu is not inventoried in Guthrie (1948, 1953, 1971) or Maho (2003, 2009) and this is why we assigned to it the code B80x. There is no glottocode or ISO 639-3 code for Boma Nkuu in the Glottolog (Hammarström *et al.* 2018) or the Ethnologue (Eberhard *et al.* 2019).

The fourth group inventoried by Tonnoir (1970: 4) is called *(Ba-)boma (Ba-)nkuu sud* and is located just southwest with respect to the neighboring Boma Nkuu, in the current DRC province of Kinshasa (see B80y in Map 2). This group is reported in a map in Sulzmann (1983: 532) and Stappers (1986: VI). In both cases, it is labeled in German *(Ba-)Boma süd*. The geographical location of this group in these two sources is different from the one in Tonnoir (1970). In some sources, presumably both the third and fourth group discussed in this section are called *boma nkim* (Roelants 1969). We named this variety South Boma [Nkuu] and assigned to it the code B80y, following the same reasoning outlined above for Boma Nkuu B80x and Boma Yumu B80z. To our knowledge, the only data on South Boma [Nkuu]

is a wordlist in Nsuka Nkutsi (1990), who labels this variety *boma*.[9] Judging by the little available data, South Boma [Nkuu] appears to be distinct from all other Boma languages discussed here. There is no glottocode or ISO 639-3 code for South Boma [Nkuu] in the Glottolog (Hammarström *et al.* 2018) or the Ethnologue (Eberhard *et al.* 2019).

Our hypotheses on the linguistic status of these "Boma" language varieties are supported by the phylogenetic tree discussed in Section 3. As shown in Figure 1 in Section 3.3, all four Boma varieties displayed in Map 2 are part of WCB's major subclade Kasai-Ngounie Extended, but they do not cluster together as a distinct subgroup within that subclade, quite the contrary. Boma Yumu B80z and North Boma B82 are the only varieties that belong to the same genealogical subgroup within Kasai-Ngounie Extended, i.e. Kwa-Kasai North.

Finally, we note that since the late 19th century there has been a tendency to consider people speaking Boma B74, also known as Eboo or Eboo-Nzikou (cf. Raharimanantsoa 2012a, b; 2017 for details) in the Republic of the Congo and people speaking North Boma B82 in the DRC as the same ethnic group (Johnston 1883: 707; 1908: 881). Perhaps as a consequence of this assumption, some authors also assume that Boma B74 and North Boma B82 are the same language (Stappers 1986: V; Koni Muluwa & Bostoen 2015: 14). However, based on a comparison of their synchronic sound systems, their historical sound changes and their "basic vocabulary" items, none of the Boma varieties spoken in the DRC and discussed in this section can be considered as the same language as, or as a regiolectal variety of, Boma B74 spoken in the Republic of the Congo. This observation is again borne out by the phylogenetic classification presented in Section 3. Boma B74 is also part of the Kasai-Ngounie Extended subclade, but does not closely cluster together with any of the Boma varieties discussed in this section.

To minimize the confusion surrounding "Boma" glossonyms, we re-named B74 Eboo-Nzikou instead of B74 Boma. We believe that the conflation of North Boma (B82) spoken in the DRC with Eboo-Nzikou (B74) spoken in the Republic of the Congo possibly arose due to the fact that Eboo (B74) has *Boma* as an alternative name, an exonym used by Kongo speakers to refer to this Teke group (see also Vansina 1966: 103). This alternative name is commonly used by Bantuists and Niger-Congo scholars more generally (cf. for instance Guthrie 1970: 12; Guthrie 1971: 37; Welmers 1971: 785). In earlier editions of the Ethnologue (Grimes 1992: 233, 432; 2000: 77, 103), a language called Boma (with alternative names Boõ, Eboo, Eboom, Boma Mbali, Bamboma) belonging to the Central Teke group is reported as spoken both in the Congo and the DRC. In more recent editions (Lewis *et al.* 2016: 117), a language called Teke-Eboo (with alternative names Aboo, Bamboma, Boma, Boo, Boõ, Central Teke, Eboo Teke, Eboom, Iboo, Teke-Boma) is reported as being indigenous to the Congo and the DRC. The maps in different Ethnologue editions (Grimes 2000: 24; Lewis *et al.* 2016: 388) clearly show that what we call North Boma B82 is subsumed under "Central Teke". In the Ethnologue maps in the

9. We are relatively confident that the *boma* wordlist in Nsuka Nkutsi (1990) is South Boma [Nkuu] because the place names for this variety are located in the area that previous scholars identify as *boma nkuu sud* (Tonnoir 1970) or *Boma süd* (Sulzmann 1983; Stappers 1986).

aforementioned editions, Central Teke starts in the northeastern part of the Republic of the Congo and extends over the Congo River into the northwestern part of the DRC, where it has Tiene B81 as a neighbor. However, as discussed above, this appears to be a misinterpretation. The neighbors of Tiene B81 in the DRC are, among others, North Boma B82 and Mosieno B76a (see Map 1). We do not have access to the maps in the latest edition of the Ethnologue (Eberhard *et al.* 2019), but judging by the information under "Teke-Eboo", the situation appears to be the same as in previous editions. In the Glottolog (Hammarström *et al.* 2018), Eboo-Nzikou (B74) is called Teke-Eboo-Nzikou and is assigned the glottocode [teke 1278] and the ISO 639-3 code [ebo]. The Glottolog (Hammarström *et al.* 2018) erroneously states that this variety is spoken both in the Republic of the Congo and the DRC. The geographical coordinates for Teke-Eboo-Nzikou (-2,62; 17.03) are misleading because they point to a place in the DRC where North Boma (B82) is spoken (cf. supra).

2.3. Mpuono, Mpuun and/or Mbuun

Another labeling conundrum we address has to do with the glossonyms Mpuono, Mpuun and Mbuun. To disentangle this situation, we need to consider changes in language names for the so-called Tende-Yanzi group in Guthrie (1948, 1953, 1970). These are reproduced in Table 3 in the Appendix, which also includes the updates of Maho (2009) for this same group.

In the three successive referential classifications of Guthrie (1948, 1953, 1970), there seem to be correspondences among glossonyms throughout the years despite occasional reordering (cf. Mfinu and Tiene in Guthrie 1948 vs. Guthrie 1953 and 1970). It is highly likely that Mbunu 48 in Guthrie (1948) corresponds to Mbuun B88 in Guthrie (1953) and Mbuun B87 in Guthrie (1970). As Guthrie (1948: 81) himself points out, this language is "spoken in Congo Belge to the east of the R[iver] Kwilu in the region of Kikwit and as far east as the R[iver] Longe". The map in Guthrie (1953) confirms the location of Mbuun east of the Kwilu River in the DRC. The linguistic documentation on Mbuun is quite copious compared to other languages in Guthrie's B80 group (Dibata Mimpiya 1977; Mundeke 1977, 1979; Yome Aya 1997; Mwense 2000; Mundeke 2006; Bostoen & Mundeke 2011a, b; Mundeke 2011; Bostoen & Mundeke 2012, among others). In all linguistic sources we consulted, the alphanumeric code assigned to Mbuun is B87, probably following Guthrie (1970, 1971). In ethnographic and anthropological studies, Mbuun speakers are often named *Bambunda* (Struyf 1931; De Decker 1942), *Babunda* (Torday and Joyce 1907; Torday 1919; Torday & Joyce 1922; Flament 1934; Maes 1934; van Bulck 1948 amongst others) or *Ambundu* (Weekx 1937a, b). Mbuun speakers should not be confused with Mbunda (K15) speakers.

Unlike with Mbuun, information on Mpuon(o) and Mpuun(o) is very scanty. Guthrie (1971: 38) says that what he calls Mpuon B84a and Mpuun B84b are "broadly similar" to Mfinu B83 in terms of phonological features. The glossonym "Mpuono" appears in Johnston (1919) but is absent from major linguistic and ethnographic surveys of the DRC such as van Bulck (1948) and Boone (1973). In a recent linguistic atlas of the DRC (Maalu-Bungi *et al*. 2011: 105), Mpuono is

reported among the DRC Bantu languages for which there is no documentation. There are wordlists in Bastin *et al*. (1999) for two varieties of Mpuono B84 (cf. #56 and #57), one spoken just east of Kinshasa and another in the vast region northeast of Kinshasa and south of the Kwa River (see Map 1). Judging from the wordlists, Mpuono B84 is drastically different from Mbuun B87 and resembles some so-called "South Teke" varieties such as Fumu B77 and Wuumu B78 (see also Hammarström 2019: 26). This is further confirmed by the phylogenetic classification presented in Section 3.

As can be seen in Table 3, in Maho's (2009) B80 Tiene-Yanzi group, the code B87 has been eliminated and Mbuun is subsumed under B84 Mpuono (cf. also Grimes 2000: 98 where Mbuun is an alternative name for Mpuono). Based on the immediately preceding discussion, we believe that this collapsing of two different alphanumeric codes is misleading and should be amended to avoid proliferation of confusion. Mpuono and Mpuun should be labeled B84 and Mbuun should be kept separate under B87 (see Hammarström 2019: 26). In the Glottolog (Hammarström *et al*. 2018), the glottocode mpuo1241 and the ISO 639-3 code [zmp] identify a single language named Mbuun with alternative names Ambunu, B84_Mbunda, Babunda, Bambunda, Embuun, Mboon, Mbun and Mpuono. The dot on the map for this language corresponds to one of the locations where Mbuun B87 is spoken. As far as we can tell, Babunda, Bambunda, Embuun, Mboon, Mbun can be considered as alternative names for Mbuun B87. Mpuono on the other hand appears to be a different (undescribed) language and should have a different alphanumeric code (i.e. B84) and a different geographical location compared to Mbuun B87. This is borne out by our new phylogenetic classification presented in Section 3. Mpuono B84 and Mbuun B87 clearly belong to two distinct genealogical subgroups within WCB, i.e. Kwilu-Ngounie and KLC extended respectively.

Finally, the glossonyms Mpuon(o), Mpuun(o) and Mbuun should not be confused or taken as synonyms of Mbuno (a.k.a. Kimbuno), which appears to be a Teke variety spoken in the DRC (Johnston 1919; Bittremieux 1936; van Bulck 1948: 492).

2.4. Teke, Tyo and/or Tio

In this section, we address some referential classifications of so-called Teke languages which (at least partially) contributed to the confusing use of the terms Teke and/or Tyo, also spelled Tio (see Table 3 in the Appendix). These classifications include the misleading labeling of Teke varieties based on cardinal points (Guthrie 1970; Grimes 1992, 2000; Maho 2009). In general, the terms Teke and/or Tyo have been used since colonial times by different authors to mean different things or refer to different entities, sometimes peoples, sometimes (a group of) languages, without always being properly defined. Thus "Teke" does not mean anything if one does not associate the use of this exonym to a particular author and their definition, which is unfortunately lacking in most instances.

Teke is a Kongo exonym for "all the populations who live on the plateaus north of Malebo Pool on both sides of the Congo River as far north as the mouth of the Nkeni" (Vansina 1966: 102), an affluent of the Congo River located roughly at

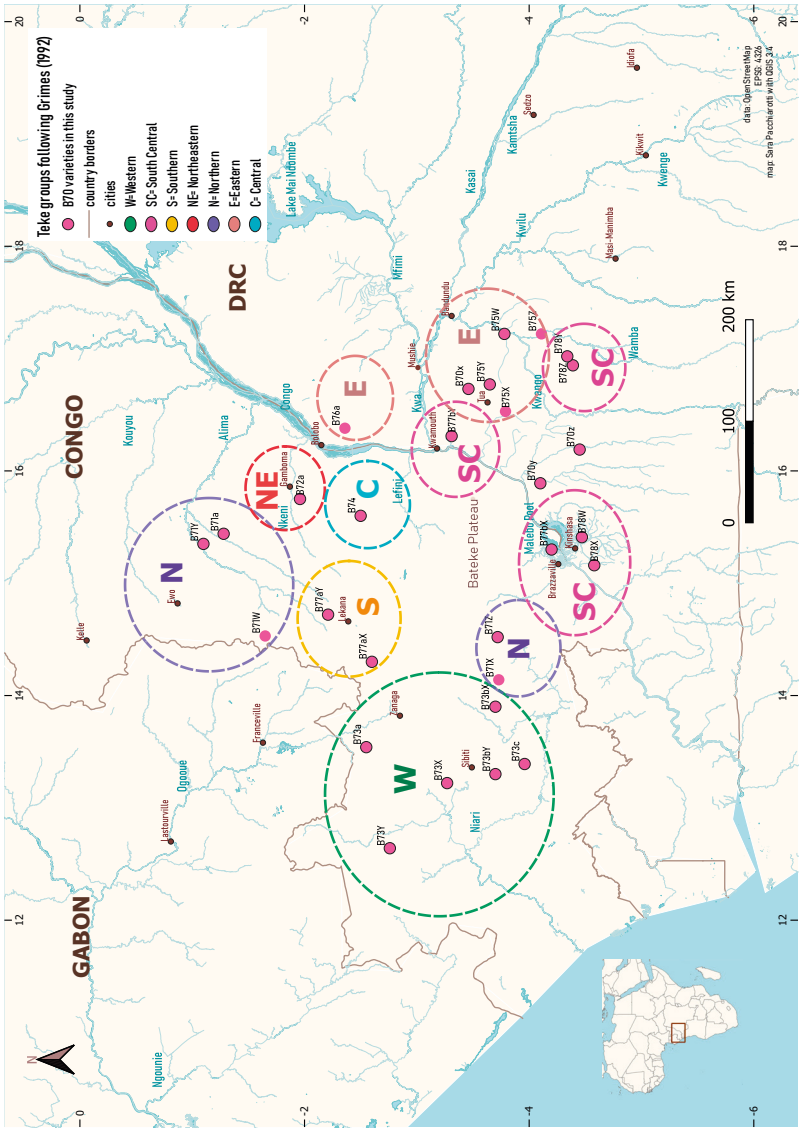
16.30, -2.02 in the Republic of the Congo (see Map 1). Teke peoples designate themselves as *tyo* or *teo* (Jacquot 1965: 340; Vansina 1966: 102; Boone 1973: 295 and ff. and references therein). Usually Teke peoples name their subgroups after the plateau where they live. Vansina (1966) notes that populations who are "culturally related" to the Teke but do not live on the plateaus are also sometimes called Teke (cf. Nsuka Nkutsi 1990 for an example of this usage). Apparently, Teke has also been used to refer to varieties somehow linguistically related to Teke languages (see for instance Adam 1954: 34).[10]

At the present time, we do not have a good understanding of the internal structure of the so-called Teke group nor do we know whether it should be considered as a dialect continuum. We refer the interested reader to a few available dialectological and lexicostatistical studies, such as Bouka (1989), Ndamba (1994, 1996) and Bouka & Ndamba (1992). Several others exist according to some sources (Raharimanantsoa 2012a; Idiata *et al.* 2013), but they are unpublished and/or not easily accessible (Kristensen *et al.* 1984; Linton 1999). We note that there are several Teke varieties that are not referenced in Guthrie (1971) or Maho (2009) nor included in this study. Some of these include Teke Zanaga (Kristensen *et al.* 1984; Raharimanantsoa 2012a) and Ntsi-ntsege (Jacquot 1971; Kristensen *et al.* 1984). A map of the current distribution of Teke varieties in the Republic of the Congo can be found in Kristensen *et al.* (1984: 10). Additionally, although not included in this study, we note that a Teke variety called Lateghe spoken in southeastern Gabon has been documented in detail by Linton (2009, 2013a, b; 2014, 2016). Linton places Lateghe under B71b together with Kateghe.

Besides the idiosyncratic use of the terms Teke and/or Tyo/Tio, labeling across different referential (i.e. geographical) classifications further aggravates the confusion. We illustrate this great deal of confusion with some referential classifications of so-called Teke languages spoken both in the Republic of the Congo and the DRC in Table 4 in the Appendix. In Table 4, each column represents a Teke referential, non-genetic subgrouping according to a given author. For the sake of clarity, we disrupted consecutive numbering of Teke varieties in Guthrie (1948) so that their names would match those of Guthrie (1953), cf. for instance "31 Fumu" in Guthrie (1948) which we placed in the cell next to "B78 Fumu, i-" in Guthrie (1953). Alphanumeric codes assigned to Teke varieties in Guthrie (1953) were substantially reshuffled in Guthrie (1970), the source on which virtually all

10. "La famille des langues Téké est surtout répandue au Moyen-Congo et au Congo belge. Cependant au moins de ¼ de la population du Gabon parle des dialectes apparentés au Téké." [The Téké language family is spread through Moyen-Congo [read: the Republic of the Congo] and the Belgian Congo [read: DRC]. Nevertheless, at least ¼ of the population of Gabon speaks dialects related to Téké] (our own translation). Elsewhere in the literature, B60 languages are said to "show similarities" with the Teke group (see Idiata *et al.* 2013: 186 and references therein). Lane (1989) reports that some official policy makers in the Republic of the Congo consider Nzebi (B50) and Mbete (B60) to be dialects of Teke. However, "Mbete people refuse to say that they are Teke, but accept Njebi [read Nzebi] as a subgroup of Mbete. Yet, the Teke say that they are the same people as the Mbete or Njebi" (Lane 1989: 13). For a detailed account of the multilingualism, sociolinguistic dynamics and linguistic attitudes of Teke (B70) speakers and Mbete and Mbaama (B60) speakers in northern and southern Republic of the Congo, see Lane (1989: 29 and ff.).

subsequent referential classifications are (at least partially) based (Jacquot 1971; Grimes 1992, 2000; Maho 2003, 2009; Lewis *et al.* 2016). It is clear from Table 4 that Teke, Tyo and their alternative spellings have been used to refer to both groupings and individual languages, see for instance Guthrie (1970) W. Teke B73, and Teke as an alternative spelling of Tio B75. In some instances, however, glossonyms that look like alternative spellings seem to refer to distinct entities, cf. Tyoo in Grimes (2000) vs. Tio in Guthrie (1953, 1970) and Maho (2009) and Tyo in Grimes (2000). We would like to draw the reader's attention to the fact that prior to Grimes (1992, 2000), the first attempt to label some Teke varieties based on cardinal points is presumably Guthrie (1953), cf. B71 N[orth]. Tege and B77 S[outh].W[est]. Tege in Table 4. This attempt is fully elaborated in Guthrie (1970) and maintained in Maho (2009). Grimes (1992, 2000) also follows the Teke cardinal-point-based referential classification of Guthrie (1970), but with two differences. First, Grimes (1992, 2000) subsumes Guthrie's (1970) Bali (Tio, Teke) (B75) and East Teke (B76), which includes Mosieno (B76a) and Ŋgee (B76b), under "Eastern Teke" (see Table 4). Possibly because of this re-grouping, some authors label Tio (B75) as B76 (cf. Bastin *et al.* 1999: 12 ; Koni Muluwa & Bostoen 2015: 30), even though B76 varieties such as Ŋgee (B76b) are likely to be different languages from Tio (B75) (Lewis *et al.* 2016: 139). Second, Grimes (1992, 2000) adds Guthrie's Fuumu (B77b) to Wuumu (Wumbu) (B78) and labels these two "South Central Teke" (see Table 4). Unfortunately, the cardinal points used by Guthrie (1970) and Grimes (1992, 2000) to refer to different Teke groups do not seem to have a purely geographical basis and are misleading. This is perhaps why Lewis *et al.* (2016) depart from cardinal-point-based labeling (see Table 4). We show the mismatch between the labeling of Teke groups based on cardinal points and their actual location in Map 3. In Map 3, we group the Teke varieties in our dataset (which are unfortunately named following the cardinal points format, e.g. North Teke) by assigning different colors to each of Grimes' (1992, 2000) cardinal-point-based group (see Table 4 in the Appendix).



Map 3. B70 varieties in our dataset grouped according to the Ethnologue (Grimes 1992, 2000)

As can be seen in Map 3, the only labels which somehow reflect the actual geographical location of the corresponding varieties are Western and Eastern Teke. A striking example of the mismatch between the labels used in cardinal-point-based groups and the actual geographical location of varieties involved are the North Teke B71 varieties circled in purple in Map 3. Bastin *et al.* (1999) have North Teke B71 wordlists from four different departments in the Republic of the Congo: Plateaux, Cuvette, Bouenza and Pool. The upper purple circle in Map 3 contains B71W (Cuvette), B71Y (Cuvette) and B71a (Plateaux), while the lower purple circle contains B71X (Bouenza) and B71Z (Pool). Note that the lower purple circle is located, however, just *south* of the yellow circle containing so-called South Teke varieties (Kukwa B77aX and Kukwa B77aY). South-Central Teke is also a problematic label. Based on the geographical location of Fumu (B77b) and Wuumu (B78) varieties in our dataset, we get three pink South-Central circles in Map 3. The pink circle containing the Fumu variety (B77bY) is located roughly at the same height as the lower purple circle containing North Teke varieties, but eastwards. A second pink circle containing the Wuumu varieties (B78X) and (B78Y) is located just below one of the brown circles containing some Eastern Teke varieties. This obviously creates a clash between Eastern and Central labels. Although the classification of Lewis *et al.* (2016) and the latest edition of the Ethnologue (Eberhard *et al.* 2019) refrain from using cardinal points to name Teke varieties, most Bantu scholars use Maho (2009) as a reference model, which is entirely based on Guthrie (1970) and to a great extent reflects the referential classification of Grimes (1992, 2000), as illustrated in Map 3.

We conclude this section by highlighting that our understanding of the internal genealogical relations among so-called Teke languages is still very limited. Our phylogenetic study in Section 3 shows that Teke varieties labeled with the same glossonym show up in different places within our tree. This is true especially of so-called North Teke (B71), Fumu (B77b), Tio Bali (B75) and Wuumu (B78). This means that in principle these doculects are *not* different instantiations of one and the same language even though the glossonyms tell us so. It might well be the case that doculect labels reflect membership other than linguistic (e.g. ethnic) and that there might be more distinct varieties of Teke than the ones we know.

2.5. Mfununga, Mfunika, Mfunuka, etc.

Speakers of languages located in the area northeast of Kinshasa delimited to the north by the Kwa River, to the east by the Kwango River and to the west by the Congo River (see Map 1) are often called in the literature *Fumu, Mfumu, Nfumuka, Mfumungu, Mfunika, Mfono, Mpfuono, Mfunu, Nfungumu, Mfungunu, Mfunuka, Mfununga, and Mfunungu* among others (de Vos 1910: 87; Maes 1924: 7; Bittremieux 1936: 665; van Bulck 1948: 488; 1954: 63; Boone 1973: 229, 237). The use of these names for populations inhabiting this area can be seen for instance in ethnolinguistic maps of the DRC in Soret (1955), Vansina (1966: 131) and Sulzmann (1983: 525). All of these terms are exonyms originating in the (presumably) vehicular Kongo verb **ku-funu** 'to gather, to join together' (Boone 1973: 229) and mean something akin to 'the ones who gathered and mixed with others.' The exonym *(Ba-)mfununga* (and variations thereof as specified above) was used by Kongo speakers and

Europeans during colonial times to refer to *any* non-Kongo peoples inhabiting the area northeast of Kinshasa delimited by the Congo River to the west, the Kwango River to the east and the Kwa River to the north (de Vos 1910: 87, see also the discussion in Sülzmann 1983: 529).

There are at least three presumably distinct languages which are often referred to with the exonyms mentioned in the preceding paragraph. These are Fumu (B77b), Wuumu (B78), and Mfinu (B83). See Maalu-Bungi *et al.* (2011: 24) for several other (non-)Teke varieties in the DRC which have Mfununga as an alternative name.

Fumu (B77b) (see #36-37 in Table 2) is referentially classified as a Teke variety (cf. Table 3) spoken around Malebo Pool (formerly known as Stanley Pool) to the north of Brazzaville in the Republic of the Congo (Jacquot 1965; Makouta-Mboukou 1976). This variety is also known as *ɨfumu* (Calloc'h 1911), *mfumu/fuumu* or *teke du Pool* (Jacquot 1965). Bastin *et al.* (1999) presumably collected Fumu data also in the DRC (see B77bY on Map 1) in the area around Kwamouth. Fumu (B77b) should not be confused with *fumu* as an alternative form of *ba-mfununga* (cf. supra).

Wuumu (B78) (see #38-41 in Table 2) is also referentially classified as a Teke variety (cf. Table 4). Alternative names include *humbu, wumu, wumbu* (van Bulck 1948; Boone 1973), *wuũũ* (Vansina 1966), *gumbu* (Masuka 1952; Boone 1973), *gum, wum, pumbu* (van Bulck 1954: 47, 93, 124) and *hum* (Vansina 1964). Wuumu (B78) appears to be spoken in several locations in the DRC (see maps in Soret 1955; Vansina 1966: 131; Boone 1973: 329; Sulzmann 1983: 532) in the area to the east of Kinshasa south of the so-called Black-River, a.k.a. Mbuampomo (de Vos 1910: 87), with pockets of speakers dispersed around both sides of the Kwango River (see B78 in Map 1). Wuumu (B78) is also spoken in the Republic of the Congo in the area north of Brazzaville, further north compared to where Fumu (B77b) speakers are located (Jacquot 1965, 1971). In Lewis *et al.* (2016: 117), Wuumu (B78) is considered as a dialect of Fumu (B77b). In Hammarström (2019: 26), Wuumu (B78) is given as an alternative name for Fumu (B77b).

Mfinu (B83) is spoken east of Kinshasa in the DRC (cf. Map 1) but is not usually classified, at least from a referential standpoint, as a Teke variety. Nevertheless, Mfinu (B83) speakers are also often referred to with one of the alternative names discussed in this section, *Bamfununga, Munika* and *Mfono* (cf. Table 2, Guthrie 1953: 81; 1956: 84; Boone 1973: 229).

3. A new phylogenetic classification of the B50-80 languages

In this section, we offer a new lexicon-based phylogeny of the B50-80 languages generated from the largest possible sample of doculects available to us. We present and discuss the new phylogenetic tree in Section 3.3. In Section 3.1, we present a short overview of the most important previous quantitative lexicon-based approaches to the internal classification of WCB. In Section 3.2, we introduce the phylogenetic approach used to generate this tree and we explain how we went about cognacy judgments.

3.1. Previous quantitative approaches to WCB classification

This section discusses major previous quantitative approaches to internal WCB classification, both lexicostatistical (Bastin *et al.* 1999) in Section 3.1.1, and phylogenetic (de Schryver *et al.* 2015; Grollemund *et al.* 2015; Bostoen & de Schryver 2018a, b) in Section 3.1.2. We summarize commonalities and differences between these two approaches in Section 3.1.3.

3.1.1. Bantu B50-80 in the lexicostatistical study of Bastin *et al*. (1999)

As for most of the world's language families, lexicostatistics was for a long time the main quantitative approach to the genealogical classification of Bantu languages. This method calculates the degree of similarity among related languages on the basis of a limited set of so-called "basic vocabulary". This part of the lexicon is assumed to be the most stable and the least prone to borrowing, and therefore the most appropriate to reflect genealogical relations between languages. Subgroupings within a family are generated from a matrix of cognacy percentages between pairs of related languages, which reflect their degree of lexical (dis)similarity. The higher the percentage of basic vocabulary two languages share, the closer they will be related in the family tree. However, the internal structure of a given tree may vary, and sometimes quite drastically, depending on the clustering algorithm.

We will not reiterate here the details of different clustering methods within lexicostatistics, because they have been discussed in earlier studies on Bantu classification (cf. Bastin & Piron 1999; de Schryver *et al.* 2015; Philippson & Grollemund 2019). What is important to retain for our current purposes is that the last and most comprehensive lexicostatistical study of the Bantu family, i.e. Bastin *et al.* (1999), includes no less than nine different trees reflecting possible historical relations among the 542 varieties included in the study, following the nine clustering methods used in the study. These different methods use the same cognacy percentages to come to variable hierarchical subgroupings of language varieties. The cognacy percentages are based on a reduced Swadesh 100 list (cf. Swadesh 1955) of basic vocabulary (i.e. 92 items), the same we used in this study (Section 3.2). Across these nine trees, some major clusters of language varieties are robust and recurrent. This suggests that their lexical similarity is higher than that of clusters that vary from one tree to the other depending on the statistical method.

As for B50-80 languages, the core of our present study, their position within Bantu according to lexicostatistics is best captured in Table 4.2.2.2 in Bastin *et al.* (1999: 128). This table summarizes similarities and dissimilarities between the trees in terms of subgrouping. The B50-80 languages belong to a major "Western Bantu" cluster together with most languages of Guthrie's zones C, H, K (except K31), and R, groups B20 and B40 and the L21-2 languages. This Western Bantu cluster, a.k.a. "narrow West Bantu" (Vansina 1995: 186-187), is distinct from two other major clusters, i.e. "North-West Bantu" (most of zone A and B10+B30) and "East Bantu" (zones E, F, G, J, M, N, P, S and most of zones D (except D12 + D30 and L, except L2), and from several "Periphery" languages (A31, A44, A60, C43-45, D12, D30) clustering in multiple smaller subgroups (Bastin *et al.* 1999: 125).

Within Western Bantu, most of the B50-80 languages cluster most closely across the eight trees with languages of the B40, H10 (except H13b) and H30 groups and Hungan (H42), while most languages of Guthrie's zone C cluster in two distinct subgroups, i.e. "Rivers" (also incorporating B20 languages) and "Basin". Languages of zones K and R together with H13b, H43 and L21-22 form a third "South-West" subgroup.

Later lexicon-based phylogenetic studies corroborate the clustering of B50-80 with B40, H10, H30 and H42 into the "West-Western Bantu" clade (Grollemund *et al.* 2015), a.k.a. "West-Coastal Bantu" (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a, b), following the subgroup labeling proposed by Vansina (1995: 185) when previewing Bastin *et al.* (1999).

According to Bastin *et al.* (1999: 128), B40-80, H10, H30 and H42 languages further subdivide in three major clusters across the eight trees, i.e. "Kongo-Kwilu", "Nzebi", and "Teke". Kongo-Kwilu largely corresponds to what is currently known as the Kikongo Language Cluster (KLC) (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a, b). Kongo-Kwilu contains Guthrie's B40, H10, H30 groups, Hungan (H42), and one specific variety of Nsambaan, i.e. Nsambaan (Kwenge) (B85FY/[B85/7]) which is actually Samba (L12a) and has also been classified in the KLC by de Schryver *et al.* (2015) (cf. (45) in Table 2). Nzebi includes the B50-70 languages, except Tio Bali (B75) (see #29-31 in Table 2) and Wuumu (B78) (see #37-40 in Table 2). These two B70 varieties are part of the Teke subgroup together with most B80 varieties.

Finally, it is important to note that the B80 varieties in (1) have shifting alliances across the different trees in Bastin *et al.* (1999). While they cluster with the Basin subgroup in some trees, they rather go with the Teke subgroup in others.

(1) Boma Yumu (Saio) (B80zY/[B82/2]), #45 in Table 2
Tiene (DYA, Mansele) (B81X/[B81/2]), #46 in Table 2
Tiene (Bosiki) (B81Z/[B81/1]), #48 in Table 2
Mpe (B821/[B80K]), #49 in Table 2
Nunu (B822/[B80Nu]), #50 in Table 2
North Boma (Mbali-Iboma) (B82Y/[B82/1]), #52 in Table 2
North Boma (Mpukumbu) (B82Z/[B82/3]), #53 in Table 2
West Ngwi (Mateko) (B861Y/[B88/5]), #77 in Table 2
Nzadi (Panu) (B865Y/[B80Nz2]), #86 in Table 2
Nzadi (Makanga) (B865Z/[B80Nz1]), #87 in Table 2

In the same vein, certain trees attract towards the Teke subgroup languages from zones other than B, such as Kebay (C30Kb), Kesha (C30Ke1/2), Sakata (C34), certain languages from Guthrie's C80-90 groups, Mbagani (L22a), and Lwalwa (L22b). All of them are spoken in the vicinity of B80 languages, which indicates that contact-induced change may have an impact on trees generated from lexicostatistical similarity matrixes. This is also true for phylogenetic trees.

3.1.2. Bantu B50-80 in recent lexicon-based phylogenetic studies

Since the beginning of the 21st century, phylogenetic methods began to make their way into historical linguistics (cf. Dunn 2014 for an introduction to the use of phylogenetics in historical linguistics and its different methods). As a quantitative approach to language classification, they rapidly replaced lexicostatistics, even though most lexicon-based phylogenetic studies rely on similar sets of basic vocabulary and the underlying assumption that these are most resistant to borrowing. This is also the case in Bantu (cf. de Schryver *et al.* 2015; Philippson & Grollemund 2019 for recent reviews of Bantu phylogenetic studies).

So far, the most comprehensive phylogenetic classification of Bantu languages is Grollemund *et al.* (2015). This study is based on a list of 100 basic vocabulary items that largely (but not entirely) correspond to the Swadesh-100 wordlist (cf. Swadesh 1955). Out of 424 varieties in total, 32 are B50-80 languages (7 x B50, 5 x B60, 7 x B70, 13 x B80), while the KLC is represented with 30 languages (6 x B40, 21 x H10, 2 x H30, 1 x H42). Within B50-80, Grollemund *et al.* (2015) include two varieties of Duma (B51), Nzebi (B52), Tsaangi (B53), Lempini (B601), two varieties of Kaningi (B602), Mbaama (B62), Ndumu (B63), Atsitsege (B701), four varieties of North Teke (B71), Fumu (B77b), Wuumu (B78), Tiene (B81), Boma Yumu (B80z/B82), Mbuun (B87), Yans (B85), Nsong (B85d), Mpur (B85e), two varieties of Ding (B86), Ngwi (B861), Lwel (B862), Mpiin (B863), Ngong (B864) and Nzadi (B865). The data for the B50-60 and B701-B71 varieties were extracted from the database compiled for the *Atlas Linguistique du Gabon* (ALGAB) project (cf. Idiata 2007; Idiata *et al.* 2013). The data for the other B70 and the B80 varieties are the same as in Bastin *et al.* (1999), except for Tiene (B81) (Motingea Mangulu 2004), Nzadi (B865) (unpublished data from Larry Hyman), Boma Yumu (B80z/B82), Yans (B85), Mpur (B85e), one variety of Ding (B86), Ngwi (B861), and Mbuun (B87) (Burssens 1992). In other words, the doculects included in their study only partially correspond to the B50-80 doculects covered in ours.

From a sample of 100 different trees, inferred through Bayesian MCMC methods, Grollemund *et al.* (2015) draw a consensus tree from the Bayesian posterior distribution using the BayesTraits software. Within this consensus tree, manifesting variable probabilities at every node, B50-80 constitute a discrete clade together with Sakata (C34) and the B40, H10, H30 and H42 varieties included (cf. supra). Grollemund *et al.* (2015) with this clade, having a posterior probability of 0.88, as "West-Western". West-Western is sister to a clade containing both "South-Western" and "Eastern" clades. Both the node linking together South-Western (most of zones K, L, R + H21 + H321 + H41) and Eastern (most of zones E, F, G, J, M, N, P, S and some D) and the one linking this super-clade with West-Western have a low posterior probability of 0.72. This percentage suggests that internal relationships between major western Bantu subclades are not well established. Sister to the clade containing these three clades is "Central-Western" (most of zone C and some D), also with a low posterior probability of 0.74. The first major split-off is "North-Western" (comprising zone A and groups B10-30 and some Bantoid languages).

Within West-Western, the KLC (B40, H10, H30 and H42) constitutes a discrete branch with 1.00 posterior probability sister to the remainder of the clade, which constitutes itself a clade with 0.94 posterior probability. This latter clade splits into two sister clades, one having a posterior probability of 0.96 and including the B50-60-70 varieties represented in this study, and the other with a posterior probability rate of 1.00 including the B80 varieties as well as Sakata (C34). Although the lexical dataset, the language sample and the methods used differ, these three branches within West-Western Bantu as distinguished by Grollemund *et al.* (2015) roughly correspond to the Kongo-Kwilu, Nzebi and Teke from Bastin *et al.* (1999).

We end this section with the phylogenetic study by de Schryver *et al.* (2015), which focuses on the internal classification of the Kikongo Language Cluster (KLC). Out of the 95 North-Western and Western Bantu languages included, 40 are KLC languages. The remaining languages belong to the North-Western, South-Western and Central-Western clades of Grollemund *et al.* (2015) and to West-Western or West-Coastal outside the KLC. The latter are the following 26 B50-80 varieties: Wanzi (B501), Vili (B503), Duma (B51), Nzebi (B52), Tsaangi (B53), Lempini (B601), Kaningi (B602), Mbaama (B62), Ndumu (B63), North Teke (B71a), Atsitsetege (B701), Eboo-Nzikou (B74), South East Teke (B76-77), Tiene (B81), Boma Yumu (erroneously coded B82), Yans (B85, Nsong (B85d), Nsambaan (B85F), Mpur (B85e), Ding (B86), Ngwi (B861), Lwel (B862), Mpiin (B863), Ngong (B864), Nzadi (B865), Mbuun (B87). The sources for the lexical data used in this study largely correspond to those in Grollemund *et al.* (2015), i.e. ALGAB, Bastin *et al.* (1999) and Burssens (1992) as well as first-hand fieldwork data, especially for KLC varieties. Here as well, the doculects included only correspond partially to the B50-80 doculects covered in our study. However, de Schryver *et al.* (2015) do use the same reduced Swadesh-100 wordlist as we do and as Bastin *et al.* (1999) did.

The phylogenetic method used in de Schryver *et al.* (2015) is very similar to the one in Grollemund *et al.* (2015). In the Bayesian consensus tree presented by de Schryver *et al.* (2015), the KLC constitutes a discrete clade with a posterior probability rate of 0.95 and consisting of B40, H10, H30, H42 and L12a varieties. In their tree, the KLC is sister to the 15 South-Western Bantu languages included. Both the South-Western node and the one uniting the South-Western clade and the KLC have posterior probability of 1.00. The unbalanced language sample of de Schryver *et al.* (2015) accounts for the fact that the South-Western clade is drawn towards the KLC in this incomplete Bantu family tree: South-Western varieties are underrepresented compared to KLC varieties, while Eastern varieties, which form a sister clade with South-Western Bantu in Grollemund *et al.* (2015), are entirely absent. The South-Western varieties included in de Schryver *et al.* (2015) should therefore not be considered part of West-Western or WCB. However, this distortion does indicate that these two western Bantu subgroups (i.e. South-Western and West-Coastal) somehow constitute a continuum for which the cut-off point might be difficult to determine, possibly due to recent contact (cf. de Schryver *et al.* 2015: 137, footnote 26). It also shows that the way the language sample is assembled can have a significant effect on the topology of the tree in phylogenetic studies.

As for the B50-80 varieties included in de Schryver *et al.* (2015), they split in two neat clades having 1.00 Bayesian posterior probability. One clade unites all

B80 varieties except Tiene (B81) and is sister to the subgroup uniting the KLC and South-Western Bantu. The other one unites all B50-70 varieties and Tiene (B81) and is sister to the remainder of WCB, i.e. B80 minus Tiene and the KLC. Once again, although the lexical dataset, the language sample and the methods used differ, these three branches within WCB in de Schryver *et al.* (2015) roughly correspond to the Kongo-Kwilu (=KLC), "Nzebi" (B50-70 + Tiene) and Teke (B80) i, Bastin *et al.* (1999), and by extension to the three West-Western Bantu branches distinguished in Grollemund *et al.* (2015). Within the B50-70+Tiene group of de Schryver *et al.* (2015), the B50-60 varieties form, together with North Teke B71a and Atsitsege B701, a discrete clade with 1.00 posterior probability. Within that group, the B50 varieties cluster together and are sister to the cluster formed by the B60 + B71a + B701 varieties.

Finally, the phylogenetic classifications in Bostoen & de Schryver (2018a, b) differ from de Schryver *et al.* (2015) only in terms of KLC varieties. The B50-80 languages included in Bostoen & de Schryver (2018a, b) are the same as those in de Schryver *et al.* (2015) and they also cluster in the same way, though with slightly lower support values.

3.1.3. Bantu B50-80: lexicostatistics vs. phylogenetics

It is quite clear from what precedes that despite differences in statistical methods, language samples and "basic vocabulary" items used, B50-80 languages pattern in roughly the same way in quantitative approaches to Bantu internal classification since Bastin *et al.* (1999). Apart from some B80 varieties in some lexicostatistical trees generated by Bastin *et al.* (1999), B50-80 are an integral part of a major Bantu clade known as West-Coastal or West-Western. The closest relatives of that clade are Central-Western (most of zone C and some of zone D in certain studies) and South-Western (zones K and R and most of zone L + H21 + H321 + H41). Within WCB, B50-80 varieties are not part of the KLC, a.k.a Kongo-Kwilu in Bastin *et al.* (1999). They broadly split up in two distinct clades that coincide more or less with Guthrie's referential B50-70 groups (+ Tiene B81), i.e. the Nzebi group in Bastin *et al.* (1999), and Guthrie's B80, i.e. the Teke group in Bastin *et al.* (1999). It is only in Bastin *et al.* (1999), which includes the largest sample of B50-80 varieties, that the split between Guthrie's B70 and B80 is less neat, as Tio Bali (B75) and Wuumu (B78) cluster with the Teke group rather than with the Nzebi group. Neither of these B70 varieties were included in the phylogenetic studies of Grollemund *et al.* (2015) and de Schryver *et al.* (2015). Hence, a lexicon-based phylogenetic classification including the largest possible set of B50-80 varieties, as we present in Section 4.3, will reveal in all likelihood more mismatches between the referential Guthrie groups and genealogical subgrouping.

3.2. Methodology of the new phylogeny of Bantu B50-80 languages

Our dataset includes a total of 115 doculects: 97 B50-B80 doculects (basically all available sources providing us with usable data), 11 doculects belonging to the Kikongo Language Cluster (KLC), and 7 Western Bantu languages that were shown

not to be part of WCB in previous studies. The languages outside WCB following Bastin *et al.* (1999), Grollemund *et al.* (2015) and de Schryver *et al.* (2015) are: Cokwe (K11), Pende (L11) and Umbundu (R11) for the South-Western branch; Ntomba (C35a), Ngombe (C41) and Bushoong (C83) for the Central-Western branch; and Duala (A24) for the North-Western branch. As the most distantly related language, Duala (A24) served to root the tree.

Since our main interest here are the interrelationships of non-KLC WCB languages, we originally selected one language per KLC phylogenetic subgroup according to previous studies (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a, b), namely Hungan (H42) (Kongoid), Dondo (H112B) (North Kongo), Zombo (H16g) (South Kongo), Ndibu (H16b) (Central Kongo), Nkanu (H16h) (East Kongo), Cabindan Woyo (H16d) (South-West Kongo), and Punu (B43) (North-West Kongo). However, we later included three additional West Kongo varieties, i.e. Vili (H12), Yombi (H16c), and Shira (B41) because in our first trials to generate a tree, Punu (B43) was always pulled out of the KLC into a cluster with the B50 languages (see discussion in Section 3.3). After we discovered that the doculect B85/7 in Bastin *et al.* (1999) actually does not represent Nsambaan (B85F), but Samba (L12a) (cf. #45 in Table 2), it turned out that we included one more representative of the KLC's Kongoid subgroup.

Some gaps notwithstanding, each doculect consisted of a reduced Swadesh wordlist of 92 so-called basic vocabulary concepts after Bastin *et al.* (1999). In several cases more than one word is present for the same doculect and the same concept (i.e. meaning). For each doculect, we used additional sources available to us (see Table 1) to fill in gaps in the data or check Bastin *et al.* (1999)'s data when word forms seemed at odds with their corresponding meaning. All forms for each concept were assigned to cognate sets. The first and third authors did all cognacy judgments from scratch and crosschecked each other's judgments several times at different times. We used *Bantu Lexical Reconstructions* 3 (BLR3) (Bastin *et al.* 2002) to find possible proto-forms for our synchronic reflexes. To establish cognacy we relied first and foremost on the Comparative Method applied to on-going research on sound changes in Guthrie's B50-80 languages (Pacchiarotti & Bostoen 2018). We privileged correspondences of both form and meaning, but also factored in extremely common semantic shifts in Bantu such as metonymy (Bastin 1985). This approach, together with the fact that we had different doculects for the same language, allowed us to differentiate cognate sets that would have been lumped together if one relied on the principle of resemblance (Grollemund *et al.* 2015, supporting information 1 of 5) or if one had just one doculect for a given language variety. Additionally, in order to maximally diversify innovations to get as a clear picture of internal subgroups as possible, we systematically assigned, whenever relevant, different cognacy codes to reflexes which ultimately go back to one single form on a deeper Bantu level.

To illustrate these claims, consider the synchronic forms in (2) for the meaning 'arm, hand'. In the third column of (2) forms are given in phonetic transcription. When present, a slash within square brackets separates singular and plural forms. Segmentation of noun class prefixes is present only if it was already present in the original source or if we could establish it with certainty after consulting additional sources. The symbol "~" indicates forms in free alternation within the same variety.

(2) a. Nzebi (B52Y) **[xɔ́-ɔ̀xɔ̀/mí-ɔ̀xɔ̀]**
b. Tsaangi (B53) **[kɔ-ɔhɔ/my-ɔɔhɔ]**
c. Mbaama (B62X) **[kwɔɣɔ]**
d. Eboo-Nzikou (B74) **[kwɔɔ/myɔɔ]**
e. Mosieno (B76a) **[lwɔ̀, myɔ̀]**
f. Kukwa (B77aX) **[khó-òkò/myá-àkà]**
g. Boma Yumu (B80zX) **[lw-ɔ́k/mi-ɛk] ~ [lw-ak/mi-ak]**
h. Tiene (B81X) **[kɔ́-ɔkɔ/my-ɛ́kɛ]**
i. North Boma (B82X) **[k-ɔ́ːʁɔ/my-ɔ́ːʁɔ]**
j. Mpe (B821) **[e-bɔʁɔ]**
k. East Yans (B85bT) **[kwɔɔk/miɔɔk] ~ [kwéak/miéak]**
l. Mpur (B85eX) **[kɔ̂ː]**
m. East Ngwi (B861X) **[ewó]**
n. East Lwel (B862X) **[mə-lwə]**
o. Nzadi (B865X) **[lwǒ `]**
p. West Ding (B86W) **[lɔ̌, miɔ̌]**
q. West Ding (B86Y) **[lúwà/míà]**
r. Ntomba (C35a) **[lobɔ́kɔ]**

Among the available reconstructed forms in BLR3, BLR 260 ***bókò** ‘arm, hand, front paw’ looks as a suitable proto-form for (2j) and (2r). If one is led mostly by the principle of resemblance, (2a, b, c, f, g, h, i, k) could also be posited as reflexes of BLR 260 ***bókò**. In all of these cases ***b** would have been lost. Similarly, (2d) and (2l) could also be linked to BLR 260 ***bókò**, by adding loss of ***k** in C(onsonant) 2 position in addition to loss of ***b** in C(onsonant) 1 position. However, it is impossible to posit (2 a-d, f-i, k-l) as regular reflexes of BLR 260 ***bókò**. This is because an initial ***b** is never lost in the varieties listed in (2 a-d, f-i, k-l). In some of these languages, ***b** in C1 might irregularly weaken to **[w]** or **[j]** but there is no evidence ***b** > **Ø**. The forms in (2 a-d, f-i, k-l) are most likely reflexes of BLR 3451 ***joko** with reconstructed class pairings 11/4, 14/6, 15/6 (see also Paulian 1975: 82; Ellington 1977: 176; Mouele 1997: 214). In (2a, b, d, f, h, i, k) we find class pairing 15/4, while in (2g) we find class pairing 11/4. The proto-form ***joko** has a reliability value of 0 in BLR2 (Coupez *et al*. 1998). This means that the editors of BLR2 deemed the reconstruction as no longer valid (Bostoen & Bastin 2016). Nevertheless, ***joko** is the most likely proto-form to give rise to the forms in (2 a-d, f-i, k-l). Obviously, ***bókò** and ***joko** are related and likely go back to one single proto-form on a deeper level. Closely resembling proto-forms that cannot be reduced to one simple form on the basis of regular sound change are known as “osculant” reconstructions in Bantu studies (Guthrie 1962, 1967; Bostoen 2001; Ricquier & Bostoen 2008). Considering ***joko** and ***bókò** as two distinct proto-forms allows to maximally differentiate among the different reflexes we see in (2). We are now left with the reflexes in (2e) and (2m-q). At first sight these forms look quite different from others such as (2j) and (2r). If one only has the singular noun class form, as in (2m-o), or forms with a final **[a]** , as in (2q), (2e) and (2m-q) do not immediately look like cognates with (2a-d), (2f-i) and (2k-l). However, they are. In Mosieno (B76a), East Ngwi (B861), East Lwel (B862), Nzadi (B865) and West Ding (B86Y) ***k** is usually lost in C2 and ***j** is lost

in C1. The **[l]/[lw]** segment we observe in (2e) and (2n-q) is the singular noun class prefix of class 11 which pairs with noun class 4 in the plural, at least in some forms. In East Lwel (B862), the class 11 prefix has apparently been reanalyzed as part of the noun root, which is synchronically preceded by a different noun class prefix, **mə-**. Note that in some languages, the forms for 'arm, hand' undergo vowel harmony processes (cf. Tiene B81X, Kukwa B77aX) or diphthongization (East Yans B85bT, West Ding B86Y). This discussion shows that cognacy judgments for phylogenetic studies should rely first and foremost on knowledge of regular sound changes in the languages under study. Without this knowledge, forms that at first sight do not look alike can be mistakenly considered not cognate (cp. (2a) and (2m)), and forms that do look alike can be mistakenly considered (immediate) cognate (cp. (2a) and (2j)).

The above process resulted in 92 multistate characters, one for each concept, which were subsequently binary recoded, so that each binary character represents presence or absence of a cognate-concept association. This procedure resulted in a binary matrix of 1154 characters, of which 708 were parsimony-informative.

The matrix was analyzed with MrBayes 3.2.6 (Huelsenbeck & Ronquist 2001; Ronquist *et al.* 2012). We used a restriction site model for the binary characters, as it allows for asymmetry in the rates of cognate gain and loss, with a flat Dirichlet prior. The analysis was run for 500 million generations and included two independent sampling chains and six "hot" chains to improve mixing. All analyses were performed on the CIPRES Science Gateway (Miller *et al.* 2010). Convergence and burn in were assessed using MrBayes 3.2.6 and Tracer v1.7.1 (Rambaut *et al.* 2018). Majority-rule consensus trees were annotated with FigTree v.1.4.2 (Rambaut 2018).

3.3. New phylogeny of WCB B50-80 languages

Figure 1 is the Bayesian majority-rule consensus tree representing the internal phylogenetic classification of the 97 WCB B50-80 doculects considered in this study.

Figure 1 confirms that Central-Western and North-Western Bantu languages included in our study do not belong to WCB. They split off from the common trunk before the first B50-80 languages (cf. node A in Figure 1). However, this is not the case for the South-Western Bantu languages included in our study (cf. node F' in Figure 1). In contrast to earlier studies relying on a language sample covering the entire Bantu domain (Bastin *et al.* 1999; Grollemund *et al.* 2015), South-Western Bantu languages do not constitute a distinct branch in our tree, but cluster with the KLC (cf. node F in Figure 1). This behavior is also found in other partial or unbalanced phylogenetic classifications targeting the KLC branch of WCB (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a, b). We discussed this issue in Section 3.1.2 and will not further dwell on it here. It is important to highlight, however, that Figure 1 should *not* be taken as evidence for the genealogical classification of Cokwe (K11), Pende (L11) and Umbundu (R11) as part of WCB. The position of the South-Western languages in Figure 1 is an artifact of our language sample.

In this study, we considered the KLC as a discrete clade within WCB following previous studies. Although we did not constrain the topology of the tree in Figure 1 so that the KLC would come out as a monophyletic group, it did emerge as such by itself (cf. node F in Figure 1). As discussed in Section 3.2, we had originally selected one language per KLC phylogenetic subgroup according to previous studies (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a, b). However, we later included three additional West Kongo varieties, i.e. Vili (H12), Yombe (H16c) and Shira (B41), because in our first trials to generate a tree, Punu (B43) was constantly pulled out of the KLC into a cluster with the B50 languages. Adding three of Punu's closest relatives according to de Schryver *et al.* (2015) did have the desired effect, i.e. pulling Punu back into the KLC. The fact that Punu clusters with its closest neighbors from Guthrie's B50 group in the absence of its closest relatives can probably be accounted for by contact-induced change (see the discussion in Bastin and Piron 1999). This behavior suggests that language contact, amongst others, may affect the topology of phylogenetic trees, especially if they are generated from restricted and/or unbalanced language samples aimed at better understanding a specific (Bantu) subgroup.

Apart from the two issues discussed above, i.e. (1) the encapsulation of South-Western Bantu into WCB as a branch that is sister to the KLC (cf. node F' in Figure 1), and (2) the shifting alliances of Punu (B43), our phylogenetic tree corroborates the status of the KLC as a discrete clade (cf. node F in Figure 1). We will not address the internal structure of the KLC but only how it relates, as a distinct branch within WCB, to the B50-80 varieties considered here.

In the remainder of this section, we focus on the two main results of our phylogenetic study: (1) the corroboration of a second major subclade within WCB besides the KLC, i.e. a vast clade uniting all B50-70 plus B81-84 languages, which we call here Kasai-Ngounie Extended (cf. node H in Figure 1) (Section 3.3.1), and (2) the paraphyletic status of the B85-87 varieties and their relationship to a super-clade called Loange-Atlantic (cf. node B in Figure 1) (Section 3.3.2).

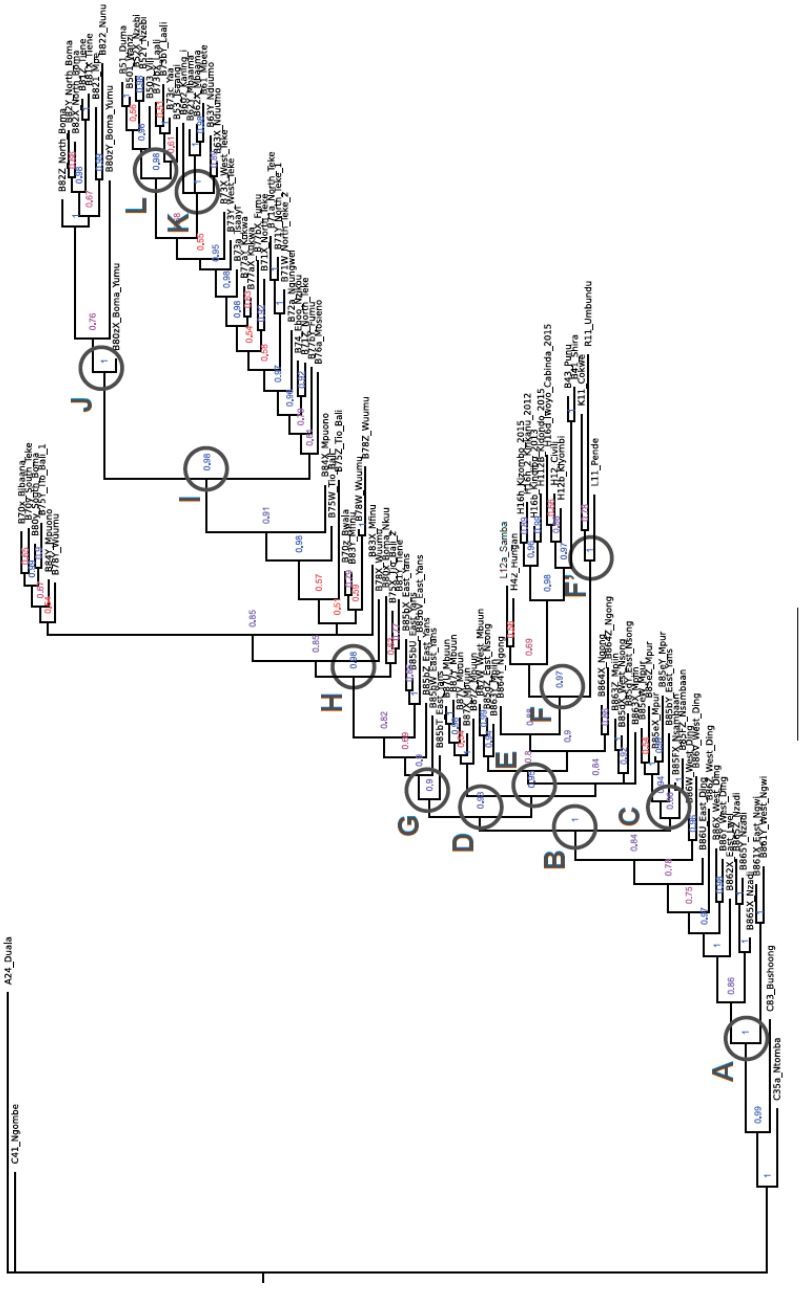


Figure 1. Internal phylogenetic classification of WCB B50-80 varieties considered in this study

3.3.1. Kasai-Ngounie Extended subclade and its internal structure

As summarized in Section 3.1.3, previous quantitative approaches to WCB classification roughly converge in distinguishing a major subgroup incorporating languages of Guthrie's B50-70 groups, to which de Schryver *et al.* (2015) add Tiene (B81) and from which Bastin *et al.* (1999) exclude Tio Bali (B75) and Wuumu (B78). In the upper part of our new phylogenetic tree we find support for such a clade. This clade includes not only all B50-70 varieties, but also all B81-B84 varieties. Unlike in Grollemund *et al.* (2015), not all B80 varieties belong to one and the same clade. Unlike in de Schryver *et al.* (2015), Tiene (B81) is not the only B80 language to cluster with the B50-70 languages. For this reason, calling this subgroup "Nzebi-Mbete-Teke" (after the names given to Guthrie's groups B50-B60-B70 respectively) is misleading, let alone the even more restrictive name "Nzebi" proposed in Bastin *et al.* (1999) following Vansina (1995). This is why we opted for Kasai-Ngounie Extended after the major water elements that delimit the distribution area of this subclade: the Kasai River to the southeast and the Ngounie River to the northwest (see Map 4). "Extended" refers to the fact that the closest relatives of the monophyletic Kasai-Ngounie group manifest paraphyly (see discussion below). The Kasai-Ngounie Extended clade is highly supported with a posterior probability of 0.98 (cf. node H in Figure 1). It includes all doculects from Guthrie's B50-70 varieties plus Boma Nkuu (B80x), South Boma [Nkuu] (B80y), Boma Yumu (B80z), Tiene (B81), North Boma (B82), Mpe (B821), Nunu (B822), Mfinu (B83) and Mpuono (B84). The clustering of the latter five varieties with B50-70 is entirely new compared to previous classifications, which either did not include them (de Schryver *et al.* 2015; Grollemund *et al.* 2015) or classified them differently (Bastin *et al.* 1999). Our understanding of the internal phylogenetic structure of the Kasai-Ngounie Extended subclade of WCB is summarized in Figure 2 and further elaborated in the following paragraphs. The tree structure in Figure 2 is represented in the form of clades in Map 4.

Within the Kasai-Ngounie Extended subclade, Boma Nkuu (B80x), South Boma [Nkuu] (B80y), Mfinu (B83), Mpuono (B84) together with Bibaana (B70x), South Teke (B70y), Bwala (B70z), Tio Bali (B75) and Wuumu (B78) split off first. Most of these languages are spoken in the DRC (except Wuumu which is also spoken in the Republic of the Congo), more specifically in an area delimited by the Kwa River to the north, the lower course of the Kwango River and the Wamba River (from its confluence with the Kwango further south) to the east and the area south of Malebo Pool to the west (see Map 4).[11] It is important to stress that these varieties do not form a clade proper or a monophyletic group, but rather what is known in phylogenetics as a paraphyletic grade. Although the section of the tree containing these doculects displays some high support values (i.e. between 0.85 and 0.98, see the area between nodes H and I in Figure 1), we consider these languages

11. Among the languages delimited by the Kwa and Wamba Rivers and the Malebo Pool, there is one odd doculect of Tiene, i.e. Tiene (Wamba) (B81Y/[B81/3]) (cf. (47) in Table 2), for which data were collected several hundred kilometers south of the language actual distribution area (see Map 1, Map 4), i.e. in Wamba to the east of Kinshasa (cf. Bastin *et al.* 1999: 11).

to have no other more recent common ancestor than the one that is ancestral to the entire KASAI-NGOUNIE EXTENDED subclade. The main reason for this is that doculects having the same glossonym, such as Tio Bali (B75), Mpuono (B84) and Wuumu (B78), end up in disparate branches within that part of the tree. This issue might be due to ethnolinguistic labeling rather than to phylogenetic classification (cf. Sections 2.3 and 2.4).

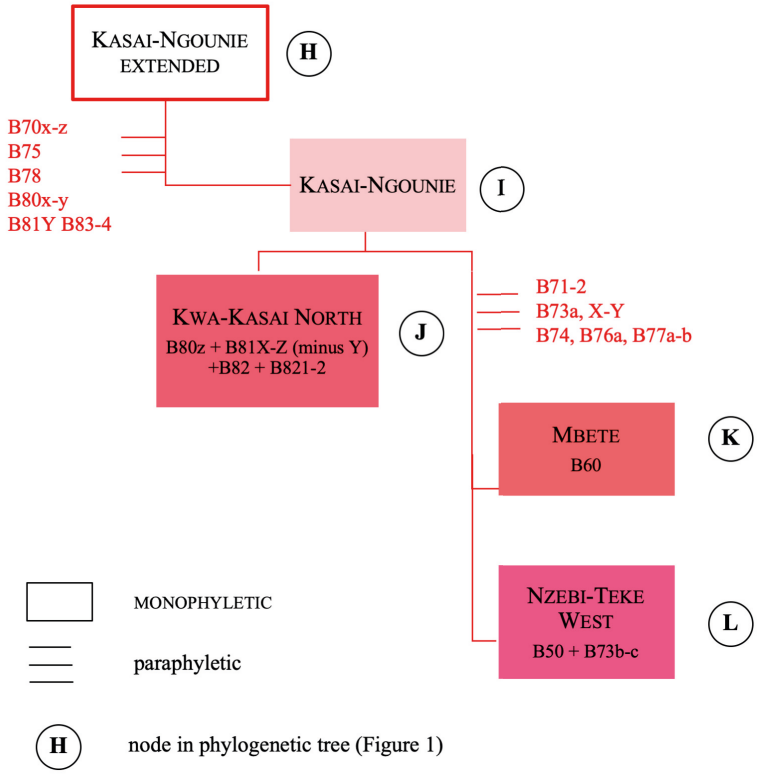
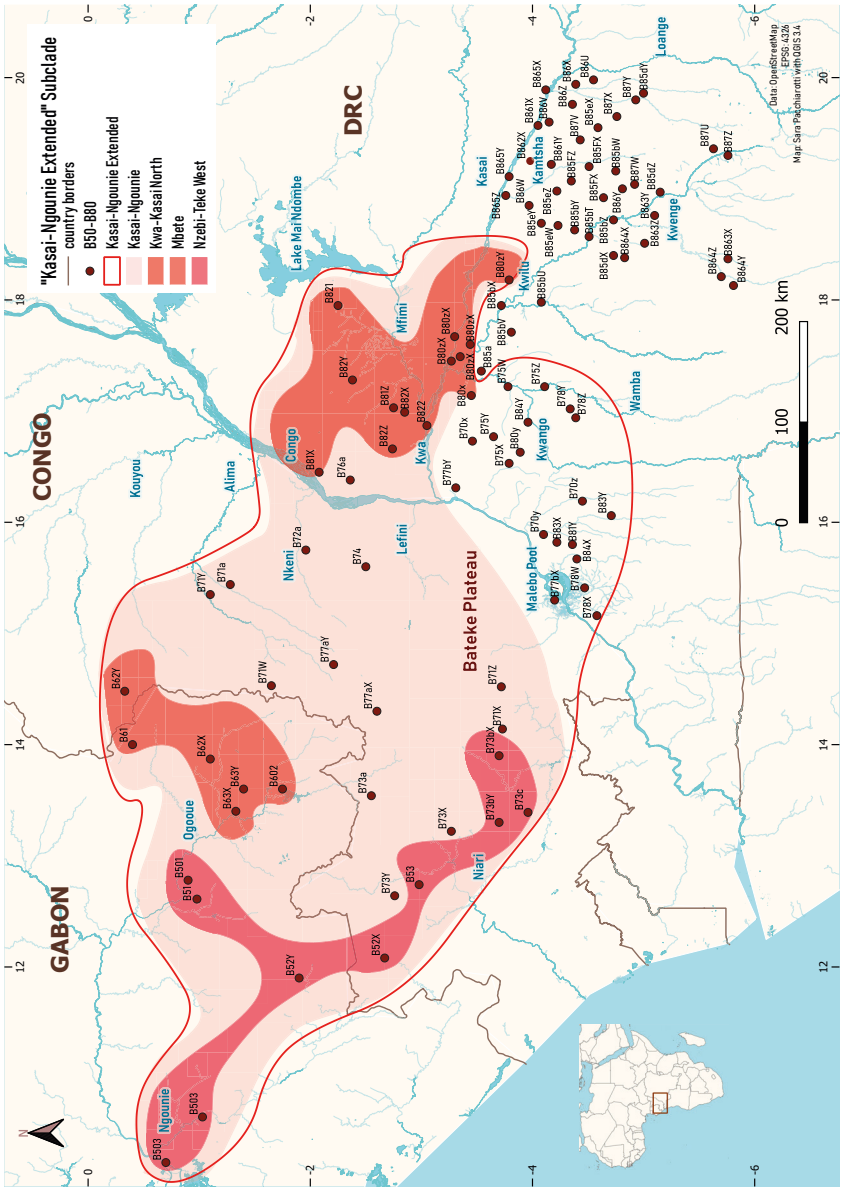


Figure 2. Internal phylogenetic structure of the KASAI-NGOUNIE EXTENDED subclade of WCB (cf. Figure 3 for its position within WCB); subgroup colors refer to Map 4



Map 4. Distribution of the KASAI-NGOUNIE EXTENDED subclade and its subgroups (cf. Figure 2)

The remainder of the KASAI-NGOUNIE EXTENDED languages do form a monophyletic group incorporating the following: B50-60; B70 minus Bibaana (B70x), South Teke (B70y), Bwala (B70z), Tio Bali (B75) and Wuumu (B78); Boma Yumu (B80z), Tiene (B81), North Boma (B82), Mpe (B821), and Nunu (B822). We call this clade KASAI-NGOUNIE, as the present-day distribution area of its member languages is confined by the Kasai River to the southeast and the Ngounie River to the northwest (see Map 4). The node representing the most recent common ancestor of the KASAI-NGOUNIE has a posterior probability of 0.98 (cf. node I in Figure 1).

The KASAI-NGOUNIE subclade itself contains some clear-cut monophyletic groups but also several paraphyletic ones. Within this subclade, the KWA-KASAI NORTH group is the monophyletic group within this subclade with the highest support value, i.e. 1.00 (cf. node J in Figure 1). This group unites the B80 varieties that did not split off earlier within the KASAI-NGOUNIE EXTENDED, i.e. Boma Yumu (B80z), Tiene (B81), North Boma (B82), Mpe (B821), and Nunu (B822).[12] Most of them are spoken north of the Kwa and Kasai Rivers, in the area delimited by Lake Mai-Ndombe to the east and the Congo River to the west. However, Boma Yumu (B80z) varieties are spoken on the southern bank of the Kasai (see Map 4).

KASAI-NGOUNIE contains a paraphyletic grade including B70 varieties spoken in between the Congo River to the east and the Ngounie River group to the west (cf. area between nodes I and K/L in Figure 1). Its highest node has a relatively low support value of 0.81.

The two clear-cut monophyletic groups within the KASAI-NGOUNIE subclade are situated much lower in the tree. One unites all B60 varieties included in this study and has a support value of 1.00 (cf. node K in Figure 1). We call it MBETE following the referential classifications of Guthrie (1948, 1953, 1970, 1971), Raponda-Walker (1960), and Jacquot (1978). The second clear-cut lower subclade with a support value of 0.98 unites all B50 varieties with Laali (B73b) and Yaa (B73c) (cf. node L in Figure 1). We call it NZEBI-TEKE WEST. Nzebi (aka Njebi, Njabi) follows the referential classifications of Guthrie (1948, 1953, 1970, 1971), Raponda-Walker (1960), and Jacquot (1978). Teke West is a cover term for Laali (B73b) and Yaa (B73c) spoken in southwestern Congo. We use Teke West instead of West Teke or Western Teke to avoid confusion with the referential classification of Teke based on cardinal points (see discussion in Section 2.4). The relatedness of B50 and some Western Teke B73 varieties is another new insight offered by this study.

The remaining paraphyletic B70 varieties within the KASAI-NGOUNIE subclade are mostly spoken on the Bateke Plateau in the Republic of Congo: North Teke (B71), Ngungwel (B72), West Teke (B73X and B73Y), Tsaayi (B73a), Eboo-Nzikou (B74), Kukwa (B77a), and Fumu (B77bX). Mosieno (B76a) and Fumu (B77bY) are situated on the left bank of the Congo River in the DRC. All these B70 varieties have a central geographical position within the wider KASAI-NGOUNIE EXTENDED subclade (see Map 4). Within our phylogenetic tree, they surface in between the ancestral node tying together the KASAI-NGOUNIE subclade (node I in Figure 1) and the lower MBETE (node K in Figure 1) and NZEBI-TEKE WEST (node L in Figure 1) subclades. Although some high Bayesian posterior probability values appear in that part of the tree, we do not consider these B70 "Bateke Plateau" languages

12. In Hammarström (2019: 26), Nunu is wrongly inventoried under B75-6.

to form a monophyletic clade for at least two reasons. First and foremost these varieties split off from the tree like the rings of an onion, maximally by clusters of two but mostly one by one, often with relatively low support values. Second, doculects presumably representing different regiolects of one and the same variety do not always cluster together. For instance, Fumu (B77bY) splits off much earlier than Fumu (B77bX). Similarly, North Teke (B71Z) splits off much earlier than the other North Teke (B71) varieties, even if it is geographically neighbor to the North Teke (B71X) doculect (see Map 4). Data for the B71X and B71Z doculects were collected in the far south of the Republic of the Congo, i.e. in the Bouenza and Pool regions respectively, while the three other B71 doculects originate further north, i.e. the Cuvette (B71W and B71Y) and Plateaux (B71a) regions. Despite the lack of a strong phylogenetic resolution among the B70 Bateke Plateau languages, it is obvious that the westernmost varieties, i.e. Kukwa (B77a), West Teke (B73), and Tsaayi (B73a), are more closely related to the Mbete and Nzebi-Teke West subclades compared to the easternmost varieties, i.e. North Teke (B71), Ngungwel (B72), Eboo (B74), Mosieno (B76a) and Fumu (B77bY).

3.3.2. Paraphyly of the B85-87 varieties and Loange-Atlantic super-clade

As discussed in Section 3.1.3, in previous quantitative approaches to WCB classification Guthrie's B80 languages cluster roughly into a discrete subgroup. However, our phylogenetic tree in Figure 1 does not support the genealogical unity of Bantu B80 languages. First, Boma Nkuu (B80x), South Boma [Nkuu] (B80y), Boma Yumu (B80z), Tiene (B81), North Boma (B82), Mpe (B821), Nunu (B822), Mfinu (B83) and Mpuono (B84) are part of the Kasai-Ngounie Extended subclade (Section 3.3.1), while the remainder of B80 languages are not. Second, most languages having the Guthrie alphanumeric code B85 or higher do not form a discrete clade elsewhere in the tree. These WCB languages belonging neither to the KLC (cf. node F in Figure 1) nor to the Kasai-Ngounie Extended (cf. node H in Figure 1) are: most East Yans (B85b) varieties, East and West Nsong (B85d), Mpur (B85e), Nsambaan (B85F), East and West Ngwi (B861), East Lwel (B862X), Mpiin (B863), Ngong (B864), Nzadi (B865), East and West Ding (B86), and Mbuun (B87). These languages are spoken in the DRC west of the Loange River, south of the Kasai River and east of the Wamba River and the lower course of the Kwango River north of its confluence with the Wamba (cf. Map 1, Map 4). Their position within the WCB tree is in Figure 3. The tree structure in Figure 3 is represented in the form of clades in Map 5.

As shown in Figure 3, some varieties do not have a more recent common ancestor than the one of the entire WCB branch. These are languages spoken along the southern bank of the Kasai River. The varieties belonging to this paraphyletic grade split off one by one from the common WCB "onion" in the following order: Ngwi (B861), Nzadi (B865), Lwel (B862), and Ding (B86), without a neat clustering of West vs. East Ding doculects (cf. area in between nodes A and B in Figure 1).

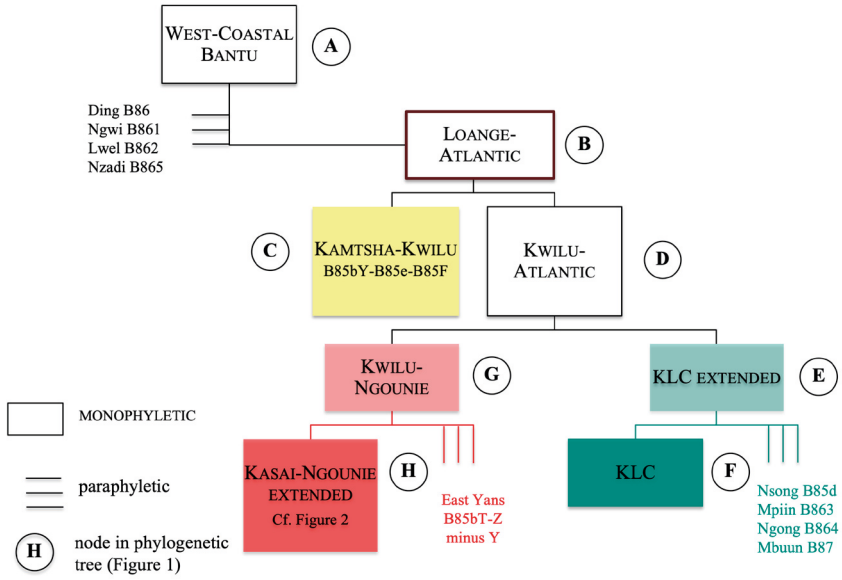


Figure 3. Internal phylogenetic structure of WCB with special reference to the B85-7 varieties; subgroup colors refer to Map 5

The remaining B85-7 varieties do belong to a monophyletic group having a most recent common ancestor that is not the same as the one to the entire WCB branch. This node with a support value of 1.00 emerges in the tree after the successive split-offs of the Kasai languages (cf. node B in Figure 1). It is a vast clade within WCB whose languages stretch from the Loange River to the east towards the Atlantic Ocean to the west. We therefore call it LOANGE-ATLANTIC. This vast clade incorporates two clades. The first is a monophyletic group (with a slightly lower support value of 0.86) comprising Mpur (B85e), Nsambaan (B85F) and one doculect of East Yans (B85bY). We name this monophyletic group KAMTSHA-KWILU after the two rivers delimiting the area where these varieties are spoken (cf. node C in Figure 1).

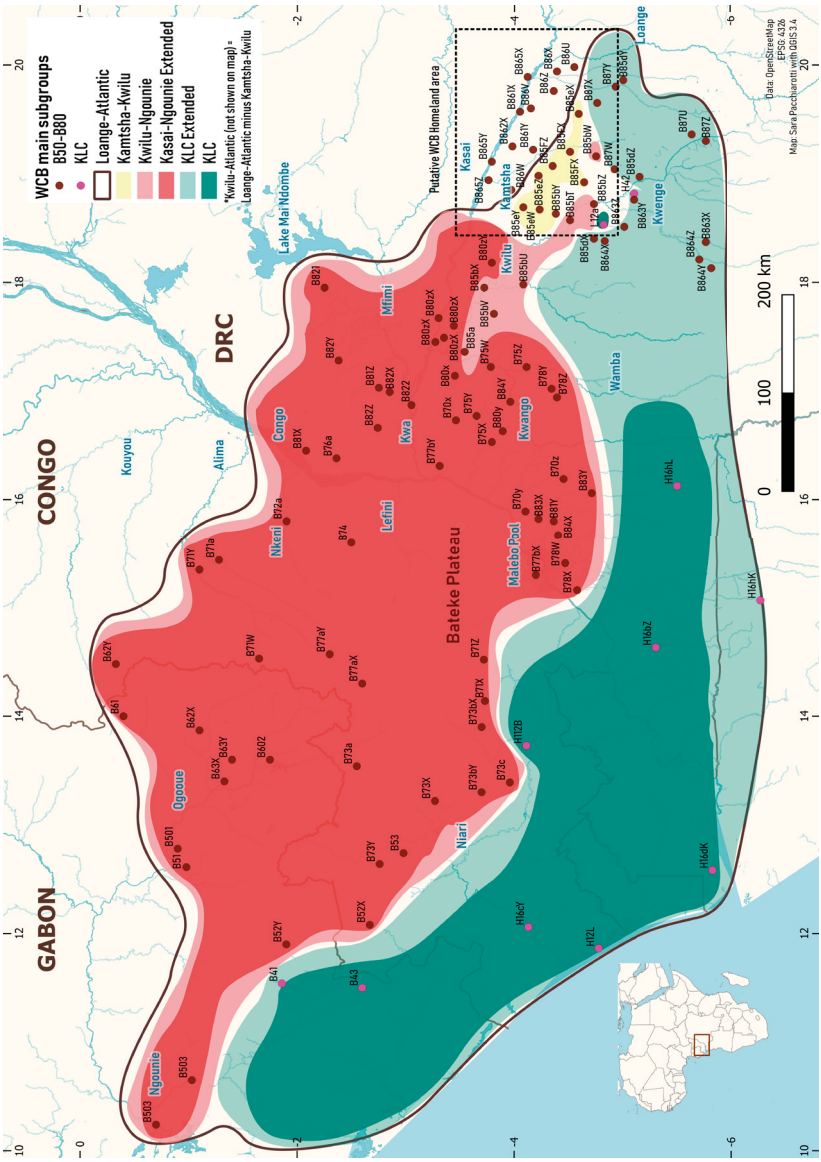
The second is the monophyletic super-cluster KWILU-ATLANTIC, which unites the KWILU-NGOUNIE and KLC EXTENDED subgroups. Its node has a support value of 0.93 (cf. node D in Figure 1). KWILU-ATLANTIC incorporates not only KWILU-NGOUNIE and KLC EXTENDED as discrete subclades, but also most East Yans (B85b) varieties, Nsong (B85d), Mpiin (B863), Ngong (B864), and Mbuun (B87) as a paraphyletic grade (cf. Figure 3). In our language sample, these latter varieties do not constitute a neat monophyletic group that could be sister to the KLC and/or the KASAI-NGOUNIE EXTENDED subclades.

East Yans doculects (except B85bY) cluster in a branch that eventually gives rise to the KASAI-NGOUNIE EXTENDED subclade (cf. area in between nodes G and H in Figure 1). As it is difficult to conceive different varieties of supposedly one and the same language as a subclade of their own, we consider them as paraphyletic. We call the monophyletic group uniting East Yans with the KASAI-NGOUNIE EXTENDED

subclade Kwilu-Ngounie after the two rivers between which its languages stretch. The node linking together the Kwilu-Ngounie subclade has a Bayesian posterior probability rate of 0.90 (cf. node G in Figure 1).

Sister to the Kwilu-Ngounie subclade is a monophyletic group supported by a Bayesian posterior probability rate of 0.95 that unites the KLC with the remaining B85-7 paraphyletic varieties, i.e. Nsong (B85d), Mpiin (B863), Ngong (B864), and Mbuun (B87). As this subclade eventually gives rise to the KLC and incorporates only four languages that do not belong to the KLC, we call it KLC extended (cf. node E in Figure 1). Among themselves, the B85-7 varieties belonging to KLC extended have no neat internal structure. As the support values associated with the nodes involving these B85-7 varieties are relatively high, the lack of phylogenetic resolution among them is unlikely to be an artifact of our lexical data or the specific language sample. The fact that doculects bearing the same glossonyms (e.g. Mpiin, Ngong, Nsong, etc. as discussed above, see areas above and below node E in Figure 1) do not cluster together possibly suggests that these labels are not based on linguistic identity. An alternative hypothesis is that some of the varieties represented by these doculects might be creoloids.[13]

13. For instance, Pokoso (1986: 7) observes that in the DRC there is not a single village made up of Ngong speakers alone. Ngong speakers are always interspersed with speakers of other varieties, such as Yans (B85), Nsong (B85d), Hungan (H42), Samba (L12a), Mbala (H41), Suku (H32) and Pende (L11). Pokoso (1986) argues that depending on the geographical location, Ngong varieties have "subjacent" varieties which heavily influence the way in which Ngong is spoken.



Map 5. Distribution of West-Coastal Bantu and its main subgroups (cf. Figure 3)

3.3.3. New understanding of the internal classification of WCB

Figure 3 not only uncovers the genealogical position of the paraphyletic B85-7 languages within WCB, but also summarizes our new understanding of how this major Bantu branch is internally organized. This new internal classification suggests that the area of earliest diversification within WCB is not situated "somewhere in between the Batéké Plateau and the Bandundu region" (Bostoen *et al.* 2015: 361).

Taking into account the present-day geographical position of the earliest split-offs, i.e. the paraphyletic varieties spoken along the Kasai (Ding B86, Ngwi B861, Lwel B862, Nzadi B865) and subsequent ramifications including B85-7 varieties, such as the Kamtsha-Kwilu clade sister to Kwilu-Atlantic within Loange-Atlantic (cf. Figure 3), the WCB homeland is rather situated in the area in between the Kamtsha and the Kasai Rivers (cf. Map 5). This is the zone of the highest diversity within WCB, where the major subclades meet. According to Occam's razor or the law of parsimony, situating the homeland in the Kamtsha-Kasai area involves the least amount of movements to account for the present-day geographic distribution of the major subclades within WCB.

It was probably somewhere within the homeland area that the Loange-Atlantic superclade diverged into the Kwilu-Ngounie and KLC extended clades, as several of their subgroups including B85-7 varieties are found today today immediately to the south-east of the Kamtsha Kasai zone, i.e. the paraphyletic East Yans (B85b) and Nsong (B85d), Mpiin (B863), Ngong (B864), and Mbuun (B87) varieties respectively.

Kwilu-Ngounie and KLC extended split off relatively early within WCB and eventually gave rise to the two monophyletic groups that underwent significant westward expansions away from the homeland, i.e. Kasai-Ngounie Extended and the KLC respectively. Kasai-Ngounie Extended spread from the homeland region all the way west up to the Ngounie River in Gabon, except for the Kwa-Kasai North subgroup that expanded north towards Lake Mai-Ndombe. The KLC expanded southwestwards roughly following the Congo River up to the Atlantic Ocean (de Schryver *et al.* 2015; Bostoen & de Schryver 2018a, b). One of its subgroups, i.e. West Kongo, spread along the Atlantic Coast towards the north where its northernmost descendants, i.e. the B40 languages, entered in contact with the westernmost Kasai-Ngounie descendants, i.e. the B50 languages. Further east, languages of the North Kongo subgroup interacted with their distant Teke relatives in the Kasai-Ngounie clade. Some varieties of the North Kongo language Bembe (H11), for instance, are known to have undergone drastic contact-induced change through intensive interactions with surrounding B70 varieties (cf. Laman 1936: LXVII; Jacquot 1962: 232).

Conclusions

In this study, we decreased the confusion surrounding glossonyms of some B60-B80 varieties and set forth necessary amendments within Bantu referential classifications. We also set the stage for future research within Guthrie's B50-80 groups. The discussion in Section 2 and the phylogenetic results in Section 3

point to the fact that (socio-)linguistic surveys and basic linguistic descriptions are needed to better understand the situation of so-called Teke languages spoken in the Republic of the Congo, the DRC and to a lesser extent in Gabon. Additionally, there are several distinct language varieties in the Republic of the Congo and the DRC in urgent need of being described. Some of these include Tsaayi (B73a), Tyee (B73d) (but see Raharimanantsoa and Ntsiba Ngolo 2015), Tio Bali (B75), Mosieno (B76a), Ng'ee (B76b), Wuumu (B78), Boma Nkuu (B80x), South Boma [Nkuu] (B80y), Boma Yumu (B80z), Mpe (B821), Nunu (B822), and Ngwi (B861).

We are well aware of the limitations of lexicon-based phylogenetic methods applied to linguistics and we believe that the study we presented in Section 3.3 is just one way of looking at the history of (a part of) WCB languages. One of the main limitations is that this study used mostly the same data as previous ones, albeit with some significant additions (cf. the lexicon in Koni Muluwa & Bostoen 2015). As observed in the introduction, we hope to compare the lexicon-based phylogenetic subgroupings in Section 3.3 with more reliable subgroupings based on shared phonological and (ideally) morphological innovations.

Despite its limitations, the phylogeny in Section 3 brings new insights on the internal classification of Guthrie's B50-80 languages and on their homeland. In terms of internal classification, our study shows that Guthrie's B50 (with the addition of some western B70 varieties), B60 and B81-84 constitute monophyletic groups. Unlike what was claimed in previous studies, the remainder of Guthrie's B70 and half of Guthrie's B80 group do not represent a genetic unit. Within Guthrie B70 and Guthrie's B85 onwards, the situation is particularly puzzling. As discussed in Sections 3.3.1 and 3.3.2, varieties in each of these two referential groups form a paraphyletic grade splitting off of a given node in the tree like rings of an onion, mostly one by one. Additionally, doculects with the same glossonyms (i.e. North Teke, Mbuun, etc.) end up in different places within the tree. As we suggested in Section 3.3.2, a possible explanation for this situation is that some glossonyms are ethnic rather than linguistic labels. Given that the support values for these paraphyletic grades are quite high, it is unlikely that the picture we get in Figure 1 is the product of chance or an artifact of the dataset. The conclusion we draw for the time being is that there seems to be more going on with these paraphyletic grades than we understand at this point in time.

Our phylogeny also points to a WCB homeland which is different from the one posited in previous studies. In contrast to previous assumptions (Bostoen *et al.* 2015: 361), the WCB homeland is not located somewhere in between the Bateke Plateau and the Bandundu region, i.e. anywhere in between 14.50 and 17.30 east, but rather further east, possibly in the area delimited by the Kamtsha and Kasai Rivers, anywhere in between 19 and 20 east, at a latitude of roughly -3.50 or southwards. We conclude by reiterating that this new WCB homeland is based on the *current* geographical distribution of WCB speakers. We do not claim that this is the only location where the homeland could be situated if we knew more about WCB population movements and the death of ancestral WCB languages it possibly involved. This new homeland hypothesis will ideally be tested against evidence from other disciplines such as archaeology.

Acknowledgements

We are grateful to Bernard Clist, Sebastian Dom, Heidi Goes, Hugh Paterson, and two anonymous reviewers for useful comments and suggestions. We are much indebted to Ruth Raharimanantsoa from SIL Congo, Barbara Thomas from Dallas International University, and Maud Devos from the Royal Museum for Central Africa for providing references we could not locate. Many thanks to Dirk Seidensticker for teaching the first author how to use QGIS and for helping her improve the maps in this paper.

References

Adam, J.J. 1954. *Grammaire composée mbede, ndumu, duma*. Montpellier: Imprimerie Charité.

Bastin, Y. 1985. *Les Relations sémantiques dans les langues bantoues*. Bruxelles: Académie Royale des Sciences d'Outre-Mer.

Bastin, Y., Coupez, A. & Mann, M. 1999. *Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study.* Tervuren: Royal Museum for Central Africa.

Bastin, Y., Coupez, A. Mumba, E. & Schadeberg, T.C. (eds) 2002. *Bantu Lexical Reconstructions 3*. Tervuren: Royal Museum for Central Africa, online database: linguistics.africamuseum.be/BLR3.html (last accessed 28 November, 2017).

Bastin, Y. & Piron, P. 1999. 'Classifications lexicostatistiques: bantou, bantou et bantoïde. De l'intérêt des 'groupes flottants'. In J.-M. Hombert & L.M. Hyman (eds), *Bantu Historical Linguistics: Theoretical and Empirical Perspectives*. Stanford: CSLI Publications, pp. 149-164.

Bissila, S. 1991. 'Description phonologique du ilaale (dialecte teke du Congo)'. Mémoire de Diplôme d'Études supérieures, Université Marien Ngouabi, Brazzaville.

Biton, A. 1969. *Dictionnaire ndumu-mbede-français et français-ndumu-mbede. Petite flore de la région de Franceville (Gabon). Grammaire ndumu-mbede*. Bar-le-Duc: Imprimerie St-Paul pour l'Archevêché de Libreville.

Bittremieux, L. 1936. 'Enkele aanteekeningen over de te weinig bekende Bateke's'. *Congo* 2 (5): 663-667.

Boone, O. 1973. *Carte ethnique de la République du Zaïre. Quart sud-ouest*. Tervuren: Musée royal de l'Afrique centrale, Series Annales in 8°.

Bostoen, K. 2001. 'Osculance in Bantu reconstructions: A case study of the pair °-kádang-/°-káng- ('fry', 'roast') and its historical implications'. *Studies in African Linguistics* 30: 121-146.

Bostoen, K. & Bastin, Y. 2016. 'Bantu Lexical Reconstruction'. In *Oxford Handbooks Online*. Oxford: Oxford University Press, http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935345.001.0001/oxfordhb-9780199935345-e-36&gt.

Bostoen, K., Clist, B., Doumenge, C., Grollemund, R., Hombert, J.-M., Koni Muluwa, J. & Maley, J. 2015. 'Middle to Late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of West Central-Africa'. *Current Anthropology* 56 (3): 354-384.

Bostoen, K. & de Schryver, G.-M. 2018a. 'Langues et évolution linguistique dans le royaume et l'aire kongo'. In B. Clist, P. de Maret & K. Bostoen (eds), *Une archéologie des provinces septentrionales du royaume Kongo*. Oxford: Archaeopress, pp. 51-55.

Bostoen, K. & de Schryver, G.-M. 2018b. 'Seventeenth-century Kikongo is not the ancestor of present-day Kikongo'. In K. Bostoen & I. Brinkman (eds), *The Kongo Kingdom: The Origins, Dynamics and Cosmopolitan Culture of an African Polity*. Cambridge: Cambridge University Press, pp. 60-102.

Bostoen, K. & Mundeke, L. 2011a. 'The causative/applicative syncretism in Mbuun (Bantu B87, DRC): semantic split or phonemic merger?' *Journal of African Languages and Linguistics* 32 (2): 179-218.

Bostoen, K. & Mundeke, L. 2011b. 'Passiveness and Inversion in Mbuun (Bantu B87, DRC)'. *Studies in Language* 21, 1: 72-111.

Bostoen, K. & L. Mundeke, L. 2012. 'Subject marking and focus in Mbuun (Bantu, B87)'. *Southern African Linguistics and Applied Language Studies* 30 (2): 139-154.

Bouka, L.Y. 1989. 'Teke and its dialects in Congo: status of the research'. In *Proceedings of the Round Table on Assuring the Feasibility of Standardization within Dialect Chains, Noordwijkerhout, The Netherlands, September 1988*, Nairobi: SIL, pp. 63-75.

Bouka, L.Y. & Ndamba, J. 1992. 'Essai de classification lexicostatistique des langues bantoues du Congo-Brazzaville'. *Afrikanistische Arbeitspapiere* 28: 33-56.

Bowern, C. 2008. *Linguistic fieldwork*. New York: Palgrave Macmillan.

Burssens, N. 1992. Vergelijkende lexicografische studie van enkele Bantutalen van de zone B. Ms. Gent (Universiteit Gent): Royal Museum for Central Africa.

Burssens, N. 1999. *Dictionnaire français-buma*. Bandundu: Centre d'Études Ethnologiques de Bandundu.

Calloc'h, J.-R. 1911. *Vocabulaire français-ifumu (batéké) précédé d'éléments de grammaire*. Paris: Geuthner.

Chambers, J.K. & Trudgill, P. 1998. *Dialectology*. Cambridge: Cambridge University Press.

Coupez, A., Bastin, Y. & Mumba, E. (eds) 1998. *Reconstructions lexicales bantoues 2 / Bantu lexical reconstructions 2 (electronic database)*. Tervuren: Musée royal de l'Afrique centrale.

Crane, T.M., Hyman, L.M. & Nsielanga Tukumu, S. 2011. *A Grammar of Nzadi (B865) A Bantu Language of the Democratic Republic of Congo with an Appendix of Proto-Bantu - Nzadi Sound Correspondences by Clara Cohen*. Berkeley: University of California Publications.

Cysouw, M. & Good, J. 2013. 'Languoid, doculect, glossonym: formalizing the notion "language" '. *Language Documentation and Conservation* 7: 331-359.

de Beaucorps, R. 1941. *Les Basongo de la Luniungu et de la Gobari*. Bruxelles: Institut royal colonial belge.

De Decker, J.M. 1942. 'Contribution à l'étude du mariage chez les Bambunda'. *Bulletin des Juridictions indigènes et du Droit coutumier congolais* 10: 125-146.

de Schryver, G.-M., Grollemund, R., Branford, S. & Bostoen, K. 2015. 'Introducing a state-of-the-art phylogenetic classification of the Kikongo Language Cluster'. *Africana Linguistica* 21: 87-162.

de Vos, S. 1910. 'Les Banfungunu'. *La Revue congolaise* 1: 87-91.

Detienne, P. 1956. *Dialectes du vicariat apostolique d'Inongo: Essai de géographie linguistique*. Louvain: Université Catholique de Louvain, Institut Africaniste.

Dibata Mimpiya, I. 1977. 'Le substantif mbuun: étude d'une forme nominale'. Travail de fin d'études de graduat, Université de Lubumbashi, Lubumbashi.

Dunn, M. 2014. 'Language phylogenies'. In C. Bowern & B. Evans (eds), *The Routledge Handbook of Historical Linguistics*. London: Routledge, pp. 190-211.

Ebalantshim Masuwan. 1980. *Esquisse grammaticale de la langue idiŋ à Káantsa*. M.A., Lubumbashi: Université Nationale du Zaïre.

Eberhard, D.M., Simons, G.F. & Fennig, C.D. (eds) 2019. *Ethnologue: Languages of the World. Twenty-second edition*. Dallas: SIL International. Online version: http://www.ethnologue.com.

Ellington, J.E. 1977. 'Aspects of the Tiene Language'. PhD thesis, University of Wisconsin, Madison.

Flament, E.B. 1934. 'Contribution à l'étude des Babunda'. *Congo* 1 (5): 679-692.

Focquet, R. & van der Kerken, G. 1924. 'Les populations indigènes des territoires de Kutu et de Nseontin (District du Lac Leopold II dans la province de l'Equateur) connues généralement sous le nom de Badia, Lesa, Basakata'. *Congo* 2 (2): 129-171.

Fontaney, V.L. 1984. 'Notes towards a description of Teke (Gabon)'. *Pholia* 1: 47-70.

Grimes, B.F. 1992. *Ethnologue (12th edition)*. Dallas: SIL International.

Grimes, B.F. 2000. *Ethnologue (14th edition)*. Dallas: SIL International.

Grollemund, R., Branford, S., Bostoen, K., Meade, A.,Venditti, C. & M. Pagel, M. 2015. 'Bantu expansion shows that habitat alters the route and pace of human dispersals'. *Proceedings of the National Academy of Sciences of the United States of America* 112 (43): 13296-13301.

Guthrie, M. 1948. *The Classification of the Bantu Languages*. London; New York: Oxford University Press for the International African Institute

Guthrie, M. 1953. *The Bantu Languages of Western Equatorial Africa* (Handbook of African languages). London; New York: Oxford University Press for the International African Institute.

Guthrie, M. 1956. 'Some features of the Mfinu verbal system'. *Bulletin of the School of Oriental and African Studies, University of London* 18, 1: 84-102.

Guthrie, M. 1962. 'A two-stage method of comparative Bantu study'. *African Language Studies* 3: 1-24.

Guthrie, M. 1967. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu languages. Volume 1: The Comparative Linguistics of the Bantu Languages*. London: Gregg International.

Guthrie, M. 1970. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu languages. Volume 3: A Catalogue of Common Bantu with Commentary*. London: Gregg International.

Guthrie, M. 1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu languages. Volume 2: Bantu Prehistory, Inventory and Indexes*. London: Gregg International.

Hammarström, H. 2019. 'An inventory of Bantu languages'. In M. Van de Velde, K. Bostoen, D. Nurse & G. Philippson (eds), *The Bantu Languages (Second Edition)*. Oxford: Routledge, pp. 17-78.

Hammarström, H., Forkel, R. & Haspelmath, M. (eds) 2018. *Glottolog 3.3*. Jena: Max Planck Institute for the Science of Human History. (Available online at http://glottolog.org last accessed on 14 January 2019).

Hochegger, H. 1972a. *Dictionnaire buma-français avec un aperçu grammatical*. Bandundu: Centre d'Études éthnologiques de Bandundu.

Hochegger, H. 1972b. *Femme, pourquoi pleures-tu? Mythes buma* Bandundu: Centre d'Études Ethnologiques de Bandundu.

Hochegger, H. 1976. *Le soleil ne se leva plus! Le conflit social dans les mythes buma*. Bandundu: Centre d'Études Ethnologiques de Bandundu.

Hombert, J.-M. & Mouele, M. 1988. 'Éléments de phonologie diachronique du wanzi (langue bantu du Gabon, groupe B50)'. *Pholia* 3: 183-205.

Huelsenbeck, J.P. & Ronquist, F. 2001. MRBAYES: 'Bayesian inference of phylogenetic trees'. *Bioinformatics* 17 (8): 754-755.

Idiata, D.F., Ratanga Atoz, A.F. & Hombert, J.-M. 2013. *Atlas des langues et peuples du Gabon*. Libreville: Les *É*ditions du CENAREST.

Idiata, D.F. 2007. *Les langues du Gabon: Données en vue de l'Élaboration d'un atlas linguistique*. Paris: L'Harmattan.

Impubi Mukwa, M. 1987. *Le syntagme nominal en yans B85d. Parler de Mikingi Kiakia - Collectivité de Nkara/Niadi*. Travail de fin d'études. Institut Supérieur Pédagogique de Kikwit, Kikwit.

Jacquot, A. 1962. 'Notes sur la phonologie du beembe (Congo)'. *Journal of African Languages* 1 (3): 232-242.

Jacquot, A. 1965. 'Précisions sur l'inventaire des langues teke du Congo'. *Cahiers d'études africaines* 5, 18: 335-340.

Jacquot, A. 1971. 'Les langues du Congo-Brazzaville. Inventaire et classification'. *Cahiers ORSTOM: Sciences humaines* 8 (4) 349-357.

Jacquot, A. 1978. 'Le Gabon'. In *Inventaire des Études Linguistiques sur les Pays d'Afrique Noire d'expression française et sur Madagascar*. Paris: CILF, pp. 493-503

Johnston, H.H. 1883. 'The River Congo, from its mouth to Bólóbó; with notes on the physical geography, natural history, resources, and political aspect of the Congo Basin'. *Proceedings of the Royal Geographical Society and Monthly Record of Geography* 5 (12): 692-711.

Johnston, H.H. 1884. *Der Kongo: Reise von seiner Mündung bis Bolobo: nebst einer Schilderung der klimatischen, naturgeschchichtlichen und ethnographischen Verhältnisse des westlichen Kongogebietes.* Leipzig: Brockhaus.

Johnston, H.H. 1908. *George Grenfell and the Congo : A History and Description of the Congo Independent State and Adjoining Districts of Congoland*. London: Hutchinson & Co.

Johnston, H.H. 1919. *A Comparative Study of the Bantu and Semi-Bantu Languages*. Oxford: Clarendon Press.

Johnston, H.H. 1922. *A Comparative Study of the Bantu and Semi-Bantu Languages. Volume II*. Oxford: Clarendon Press.

Khang Levy, N.-m.s. 1979. 'Éléments de grammaire morphologique de la langue lwel'. Mémoire de licence, Université nationale du Zaïre, Lubumbashi.

Kibwenge India'Ane, P. 1985. 'Langue mpur B85e. Phonologie et morphologie synthématique : parler de Mikingi, collectivité Dwe'. Mémoire de licence, ISP de Kikwit, Kikwit.

Koni Muluwa, J. 2010. 'Plantes, animaux et champignons en langues bantu. Étude comparée de phytonymes, zoonymes et myconymes en nsong, ngong, mpiin, mbuun et hungan (Bandundu, RD Congo)'. PhD thesis, Université libre de Bruxelles, Bruxelles.

Koni Muluwa, J. & Bostoen, K. 2015. *Lexique comparé des langues bantu du Kwilu (République démocratique du Congo)*. Cologne: Rüdiger Köppe.

Kristensen, O., Webster, D. & Webster, T. 1984. Enquête dialectale teke. Brazzaville: SIL, manuscrit non-publié.

Laman, K.E. 1936. *Dictionnaire kikongo-français, avec une étude phonétique décrivant les dialectes les plus importants de la langue dite kikongo*. Bruxelles: Librairie Falk.

Lane, B.D. 1989. 'Toward Understanding Multilingualism Among the Mbete of Northern Congo'. MA thesis, University of Texas, Arlington.

Leblanc, S., Harrison, A., Linton, P., Ngumbu, A. & Raharimanantsoa, R. 2012. *Le Développement des langues de la République du Congo: état des travaux*. Brazzaville: SIL Groupe Afrique centrale.

Lewis, M.P., Simons, G.F. & Fennig, C.D. (eds) 2016. *Ethnologue: Languages of the World, Nineteenth edition*. Dallas: SIL International. Online version: http://www.ethnologue.com.

Linton, P. 1999. Enquête dialectale teke. Ms. Brazzaville: SIL-Congo (non publié),

Linton, P. 2009. *The Noun Class System and Noun Phrases of Lategé*. Libreville: GRELACO-SIL.

Linton, P. 2013a. *La Négation en lategé*. Léconi: Global Outreach Mission.

Linton, P. 2013b. *La Phrase interrogative en lategé*. Léconi: Global Outreach Mission.

Linton, P. 2014. *Les Propositions temporelles et conditionelles du lategé*. Léconi: Global Outreach Mission.

Linton, P. 2016. *Le Système verbal de la langue lategé (deuxième revision)*. Léconi: Global Outreach Mission.

Loubelo, F. 1987. 'Description phonologique du itsaangi, parler de Madouma-Mossendjo'. Mémoire de D.E.S, Université Marien Ngouabi, Libreville.

Maalu-Bungi, C., Mutombo, H.M., Mbula, P.A., Motingea Mangulu, A., Makokila, A., Mundeke, L., Donzo Bunza, J.-P. & Nsengan, I.M. 2011. *Atlas linguistique de la République démocratique du Congo (Version révisée 2010)*. Yaoundé: Éditions du CERDOTOLA.

Maes, J. 1924. *Notes sur les populations des bassins du Kasai, de la Lukenie, et du Lac Léopold II*. Bruxelles: Van Campenhout.

Maes, J. 1934. 'Vocabulaire des populations de la région du Kasai-Lulua-Sankuru d'apres les observations de M. Achten'. *Journal de la Société des Africanistes* 4 (2): 209-268.

Maes, Y.J. & Boone, O. 1935. *Les peuplades du Congo belge: Nom et situation géographique*. Bruxelles: Veuve Monnom.

Maho, J.F. 2003. 'A classification of the Bantu languages: an update of Guthrie's referential system'. In D. Nurse & G. Philippson (eds), *The Bantu Languages*. London; New York: Routledge, pp. 639-651.

Maho, J.F. 2009. NUGL Online: The Online Version of the New Updated Guthrie List, a Referential Classification of the Bantu Languages (4 June 2009). (Online file: http://goto.glocalnet.net/mahopapers/nuglonline.pdf)

Makouta-Mboukou, J.-P. 1976. 'Étude descriptive du fumu, dialecte teke de Ngamaba, Brazzaville'. PhD thesis, Paris III, Sorbonne Nouvelle, Paris.

Marchal-Nasse, C. 1989. 'De la phonologie à la morphologie du nzebi, langue bantoue (B52) du Gabon'. PhD thesis, Université libre de Bruxelles, Bruxelles.

Masuka, A. 1952. 'Proverbes Bagumbu'. *Aequatoria* 15 (4): 135-140.

Mayanga, T. 1985. *Grammaire yansi (Rép. du Zaïre)*. Bandundu Centre d'Études Ethnologiques de Bandundu.

Mertens, J. 1939. *Les Badzing de la Kamtsha. Troisième Partie: Dictionnaire idzing-français suivi d'un aide-mémoire français-idzing*. Bruxelles: Librairie Falk fils.

Mfum-Ekong, A. 1979. 'Esquisse grammaticale yansi "parler ntsambaan" : phonologie et morphologie'. Mémoire de licence, Université nationale du Zaïre, Lubumbashi.

Mickala-Manfoubi, R. 1988. 'Éléments de description du duma, langue bantu du Gabon (B51)'. Mémoire de licence, Université libre de Bruxelles, Bruxelles.

Miller, M.A., Pfeiffer, W. & Schwartz, T. 2010. 'Creating the CIPRES science gateway for inference of large phylogenetic trees'. *Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA*, 1-8.

Motingea Mangulu, A. 2004. 'Esquisse du ketiene, petite unité du bantou B80'. *Journal of Asian and African Studies* 67: 105-150.

Mouandza, J.-D. 2001. 'Éléments de description du iyaa (parler bantu du Congo-Brazzaville)'. PhD thesis, Université de Nice Sophia Antipolis, Nice.

Mouele, M. 1997. 'Étude synchronique et diachronique des parlers dúmá (groupe bantu B50)'. PhD thesis, Université Lumière Lyon 2, Lyon.

Mouguiama-Daouda, P. 2005. *Contribution de la linguistique à l'histoire des peuples du Gabon : la méthode comparative et son application au bantu*. Paris: CNRS Éditions.

Mufanga-Dzmar, K. 1977. 'Contribution à la phonologie ding'. Travail de fin de cycle de graduat, Université nationale du Zaïre, Lubumbashi.

Mula, E. 1977. 'Esquisse de grammaire kiding munken : phonologie et morphologie (parler de Bantsiona)'. Mémoire de licence, Université nationale du Zaïre, Lubumbashi.

Mundeke, L. 1977. ‘Étude phonologique de la langue mbuun’. Travail de fin d’études de graduat, Université nationale du Zaïre, Lubumbashi.

Mundeke, L. 1979. ‘Esquisse grammaticale de la langue mbúún (parler de Éliob)’. Mémoire de licence, Université nationale du Zaïre, Lubumbashi.

Mundeke, L. 2006. ‘Éléments de phonologie et de morphologie de la langue mbuun (B87): approche structuraliste’. Mémoire de diplôme d’études spécialisées, Université de Lubumbashi.

Mundeke, L. 2011. ‘Étude morphosyntaxique de la langue mbuun (B87) (parler d’Elyob)’. Université de Lubumbashi, Lubumbashi.

Mwan Mesongolo, M. 1984. ‘Phonologie et morphologie de kedeŋ-Lesye (B86). Parler de Mateko-Mfwaay’. Mémoire de licence, Université nationale de Zaïre, Lubumbashi.

Mwense, A.M.D. 2000. ‘Étude comparative de la syntaxe mbuun (B87) et ding (B86), parlers d’Idiofa dans les secteurs Kanga et Kalanganda’. Mémoire de licence, Institut supérieur pédagogique, Kikwit.

Nsuka Nkutsi, F. 1990. ‘Note sur les parlers teke du Zaïre’. *Pholia* 5: 147-173.

Okoudowa, B. 2016. *Morfologia verbal do Lèmbáámá: Língua banta falada no Gabaõ (África Central)*. Saarbrücken: Novas Edições Acadêmicas.

Pacchiarotti, S. & Bostoen, K. 2018. ‘The internal classification of West-Coastal Bantu B50-80 languages: a preliminary assessment of diagnostic sound changes’. Paper presented at the 48th Colloquium on African Languages and Linguistics, Leiden.

Paulian, C. 1975. *Le Kukuya, langue teke du Congo: phonologie- classes nominales*. Paris: Société des Études linguistiques et anthropologiques.

Paulian, C. 1994. ‘Nasales et nasalisation en ŋgùŋgwèl, langue bantu du Congo’. *Linguistique africaine* 13: 83-129.

Philippson, G. & Grollemund, R. 2019. ‘Classifying Bantu languages’. In M. Van de Velde, K. Bostoen, D. Nurse & G. Philippson (eds), *The Bantu Languages (Second Edition)*. Oxford: Routledge, pp. 335-354.

Pokoso, K.-K. 1986. ‘Essai d’anthroponymie ngongo (parler Lukula sud)’. Mémoire de licence, Institut supérieur pédagogique, Kikwit.

Raharimanantsoa, R. 2012a. *Aspects of phonology in Eboo-Nzikou (Bantu B74)*. Preparatory research course paper, University of Gothenburg,Gothenburg.

Raharimanantsoa, R. 2012b. ‘Processes of vowel change due to second consonant loss in Boma-Nzikou (Bantu B74)’. Paper presented at the 42nd Colloquium on African Languages and Linguistics, University of Leiden.

Raharimanantsoa, R. 2017. ‘The prosody of tense marking in Teke-Eboo: A Bantu B70 language of Congo-Brazzaville’. MA thesis, Gothenburg University, Gothenburg.

Raharimanantsoa, R. & W. E. Ntsiba Ngolo. 2015. *Petit guide d’orthographe et de grammaire du téké du sud-ouest parlé dans le Pool, la Bouenza et la Lékoumou, en République du Congo (première édition)*. Brazzaville : SIL Congo.

Rambaut, A. 2018. FigTree v.1.4.4., retrieved from https://github.com/rambaut/figtree/releases.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. 2018. ‘Posterior summarization in Bayesian phylogenetics using Tracer 1.7’. *Systematic Biology* 67 (5): 901-904.

Raponda-Walker, A. 1960. *Notes d'histoire du Gabon avec une introduction, des cartes et des notes de Marcel Soret*. Montpellier: Imprimerie Charité.

Ricquier, B. & Bostoen, K. 2008. 'Resolving phonological variability in Bantu lexical reconstructions: the case of 'to bake in ashes'. *Africana Linguistica* 14: 109-150.

Roelants, F. 1969. 'Mort et funérailles dans la société boma'. In H. Hochegger (éd.), *Mort, funérailles, deuil et culte des ancêtres chez les populations du Kwango/Bas-Kwilu : rapports et compte rendu de la III^e semaine d'études ethno-pastorales, Bandundu,1967*. Bandundu: Centre d'Études ethnologiques de Bandundu, pp. 24-32.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S. Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. 2012. 'MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space'. *Systematic Biology* 61, 3: 539-542.

Rottland, F. 1977. 'Reflexes of Proto-Bantu Phonemes in Yanzi (B85)'. *Africana Linguistica* 7: 375-396.

Rurangwa, I.M. 1982. 'Éléments de description du ngungwel, langue bantoue du Congo'. Mémoire de licence spéciale, Université Libre de Bruxelles, Bruxelles.

Selvaggi, G. 1968. *La notion de dieu chez les boma*. *Dieu, idoles et sorcellerie dans la région Kwango/Bas-Kwilu: Rapports et compte rendu de la II^e Semaine d'études ethno-pastorales, Bandundu, 1966*, 67-71. Bandundu: Centre d'Études ethnologiques de Bandundu.

Soret, A. 1955. *Carte ethnique de l'Afrique Equatoriale Française : feuille n° 1 - Brazzaville*. Paris: ORSTOM.

Stappers, L. 1986. *Boma: eine Sprachskizze*. Hamburg: Helmut Buske.

Storme, M. 1956. *Ngankabe. La pretendue reine des Baboma d'apres H.-M. Stanley*. Bruxelles: Academie royale des Sciences coloniales.

Struyf, Y. 1931. 'Migrations des Bapende et des Bambunda'. *Congo* 12: 667-670.

Sulzmann, E. 1983. 'Orale Tradition and Chronologie: Der Fall Baboma - Bolia'. In C. Faik-Nzuji Madiya, E. Sulzmann, J. Daeleman & K. Ntole (eds), *Mélanges de culture et de linguistique africaines publiés à la mémoire de Leo Stappers*. Berlin: Dietrich Reimer, pp. 525-586.

Swadesh, M. 1955. 'Towards greater accuracy in lexicostatistic dating'. *International Journal of American Linguistics* 21: 121-137.

Swartenbroeckx, P. 1948. Dictionnaire kiyansi ou kiyei, langage des bayansi ou bayey de territoire de Banningville (district du Lac Léopold II) au Congo Belge. Ms. Bruxelles: Musée royal de l'Afrique centrale.

Tonnoir, R. 1970. *Giribuma : contribution à l'histoire et à la petite histoire du Congo équatorial.* Tervuren: Musée royal de l'Afrique centrale.

Torday, E. 1919. 'The Northern Babunda'. *Man* 19: 49-55.

Torday, E. & Joyce, T.A. 1907. 'On the ethnology of the south-western Congo Free State'. *Journal of the Royal Anthropological Inst. of Great Britain and Ireland* 37: 133-156.

Torday, E. & Joyce, T.A. 1922. *Notes ethnographiques sur les populations habitant les bassins du Kasai et du Kwango oriental: 1. peuplades de la forêt; 2. peuplades des prairies*. Tervuren.

Van Acht, R. 1949. 'Étude sur les devins et les jeteurs de sort chez les populations du territoire de Mushie'. *Bulletin de l'Association des anciens étudiants de l'Université coloniale de Belgique* 5: 19-47.

van Bulck, G. 1948. *Les Recherches linguistiques au Congo belge: résultats, acquis, nouvelles enquêtes à entreprendre*. Bruxelles: Librairie Falk, G. van Campenhout.

van Bulck, G. 1954. *Orthographie des noms ethniques au Congo belge suivie de la nomenclature des principales tribus et langues du Congo belge*. Bruxelles: Institut royal colonial belge.

Vansina, J. 1964. 'Noms personnels et structure sociale chez les Tyo (Teke)'. *Bulletin des Séances de l'Académie royale des Sciences d'Outre-Mer* 4: 794-804.

Vansina, J. 1966. *Kingdoms of the Savanna*. Madison: University of Wisconsin Press.

Vansina, J. 1973-1974. 'Probing the past of the Lower Kwilu peoples (Zaïre)'. *Paideuma* 19-20: 332-364.

Vansina, J. 1995. 'New linguistic evidence and the Bantu expansion'. *Journal of African History* 36 (2): 173-195.

Verhulpen, E. 1936. *Baluba et Balubaïsés du Katanga*. Anvers: Éditions de l'Avenir belge.

Vossius, I. 1666. *De Nili et aliorum fluminum origine* Den Haag: Hagae Comitis Ex Typographia Adriani Vlacq.

Weekx, G. 1937a. 'La peuplade des Ambundu (district de Kwango)'. *Congo* 1 (4): 353-373.

Weekx, G. 1937b. 'La peuplade des Ambundu (district de Kwango) (suite)'. *Congo* 2 (1): 13-35.

Welmers, W.E. 1971. 'Checklist of African language and dialect names'. In T.A. Sebeok (ed.), *Linguistics in sub-Saharan Africa*. The Hague; Paris: Mouton & Co, pp. 759-900.

Yome Aya, A. 1997. 'Étude semantique des emprunts lexicaux du francais en mbuun B87: Parler de Ntango, secteur d'Imbongo'. Mémoire de licence, ISP de Kikwit, Kikwit.

Appendix: An Inventory of West-Coastal Bantu B50-80 Varieties

Table 2 is a complete inventory of doculects used in this study. It is divided in two parts visually separated by a thick double line. The part to the left of the thick double line is organized as follows. Column "#" assigns a number to each doculect for easiness of reference. Column "Variety" lists the prefix-less names we assigned to each doculect. Parentheses after the doculect name contain the geographical reference location where a variety is spoken, e.g. #1 Wanzi (Kessipougou).[14] Slashes within the parentheses following the variety name ar used to separate two different geographical reference locations where a variety is spoken, e.g. #44 Boma Yumu (Pentane/Mondai). The geographical location of a variety might be preceded by an item in small caps, e.g. #45 Tiene (DYA, Mansele). This item represents the dialectal/regional name to which the entry belongs (cf. discussion in A.2). Column "Variety (or. source)" lists the name of each doculect exactly as we found it in the original source. Differences among glossonyms in the second and the third columns of Table 2 are discussed in A.1. Column "Code (G/M)" lists the traditional alphanumeric codes for Bantu languages found in Guthrie (1971) in their most updated form as indicated in Maho (2009). A gray cell in this column (see for instance entry #13) means that there is no alphanumeric code for a given doculect in either Guthrie (1971) or Maho (2009). Column "Code (this study)" lists the modified alphanumeric codes we assigned to each variety. The conventions we use in this column are discussed in Section 1.2. Column "Source" indicates the primary source(s) from which data were obtained. Given that Bastin *et al.* (1999) offer several distinct wordlists for the same language variety, we strived to be as precise as possible and included in square brackets the unique Guthrie-inspired code that Bastin *et al.* (1999) used for a given doculect in their lexicostatistical study, e.g. "[B51/2]" in #1.

One important thing should additionally be explained to the reader with respect to data from Bastin *et al.* (1999). We obtained these data as an Excel file transmitted by Yvonne Bastin to the last author in 2006. Not all B50-80 varieties listed in the publication of Bastin *et al.* (1999: 12-14) were present in this Excel file and not all varieties present in the Excel file were listed in the publication of Bastin *et al.* (1999). In Table 2, entries shaded in gray across the board whose original name is followed by "[not in publ. vers.]" (e.g. #62) are doculects that were present only in the Excel file we accessed but not in the published version of Bastin *et al.* (1999). To be maximally faithful to our original source, for each "gray" entry we indicate in the primary source column any additional information we found in the Excel file we obtained from Yvonne Bastin. For instance, in #62 the first name before the slash in square brackets refers to the language consultant (Fungula), while the second name after the slash refers to the transcriber (Vansina). Gray-shaded entries are further discussed in A.1.

14. In one single instance, i.e. entry #43 South Boma [Nkuu] (Boku), the doculect name includes two sets of brackets. (Boku), as in all other doculect names, refers to the location where entry #42 is spoken. [Nkuu] indicates that some sources call this variety South Boma (Nkuu) while others simply South Boma. See Section 2.2 for a detailed discussion.

The part of Table 2 to the right of the thick double line is organized as follows. Column "Coordinates (or. source)" show the latitude (N/S) and longitude values (E), respectively, for a doculect as specified in the original source. As elsewhere in this table, a gray cell in this column means that we found no coordinate information for a doculect in the original source (e.g. entry #13). Only one of our main sources, Bastin *et al.* (1999) provides coordinate values for each variety (cf. discussion in A.1). Problems and mismatches between place names and coordinate values for data from Bastin *et al.* (1999) are discussed in A.3. Column "Coordinates" shows the latitude and longitude values we provide in this study. In this column we either kept the original coordinates intact, updated them or provided them *ex novo* if lacking in the original source, see A.3. Column "Reference place" specifies a place name within a country located at the latitude and longitude values given in the column "Coordinates". The information in these three columns was obtained from the geographical database geonames.org.

To be sure, we claim neither that Table 2 exhausts *all* varieties within Guthrie's B50-80 groups nor that geographical coordinates for a particular variety represent the entire distribution area of that variety. Coordinate values in Table 2 are meant to indicate reference points exclusively.

A.1 Specifying retrievable spatial information for language varieties

All variety names in the second column of Table 2 end with a set of parentheses containing retrievable geographical information on where the variety is spoken. In a few cases, this geographical information is the same as the one found in the original source, cp. #36 Fumu (Congo) and #37 Fumu (DRC). In the vast majority of cases, however, we either added *ex novo* or modified existing geographical specifications on language varieties.

As a case of addition of spatial information, consider entry #13. This entry is named Bibaana (Dumu) instead of simply Bibaana because the original source (Nsuka Nkutsi 1990: 147) indicates that Bibaana is spoken in two main locations, Dumu and Tua, both located in the area northeast of Kinshasa and south of the Kwa River in the DRC (see Map 1). Since Nsuka Nkutsi (1990) does not state whether his Bibaana data is from Dumu or Tua (or somewhere else), we arbitrarily selected Dumu as the geographical location for the doculect named Bibaana. When a source specifically indicates that speakers of a variety come from two places, we included this information by separating the two locations with a slash, e.g. #73 Nsambaan (Kwilumpia/Longo Kuma-Kuma), after the information provided in Koni Muluwa and Bostoen (2015: 13). When two varieties were collected in exactly the same location judging by coordinate values in the original source, we added an Arabic number after the name of the variety and before the geographical information, cp. #31 Tio Bali 1 (Tua) and #30 Tio Bali 2 (Tua).[15] In general, place names in

15. We used this convention only once more, for #19 North Teke 1 (Cuvette) and #17 North Teke 2 (Cuvette). In this case, however, coordinate values for these two doculects are not identical and not even close to each other. The use of 1 and 2 here is dictated somewhat arbitrarily by the fact that both varieties were collected within the same department (i.e. Cuvette) in the Republic of the Congo.

parentheses after variety names refer to a village, town or city. Besides country names (e.g. #9-10), exceptions to this generalization are the Laali and North Teke variety names (cf. #23-24 and #16-20, respectively) which are differentiated based on department names within the Republic of the Congo, as done in the original source.[16]

In several cases, we modified the doculect geographical information found in the original source. This is especially true with data from Bastin *et al.* (1999). Let us consider as an example #92. The reader will notice that while the variety name in the original source is Ding 1 Bisey, the name we assigned to it is West Ding (KAMTSHA, Kwilumpia). Leaving aside the specifications West and KAMTSHA for now (see A.2 for discussion), the question is why we changed Bisey to Kwilumpia. This apparently unjustified change is an attempt to make the variety name maximally informative and easily retrievable in a geographical database. All varieties from Bastin *et al.* (1999) are listed in the original source with three lines of information. The first line indicates the variety name, the second information on where the variety was collected (including geocoordinates),[17] and the third names of native speakers and/or data transcribers. Thus, #92 in Table 2 appears in Bastin *et al.* (1999: 13) as in (3).

(3) B86 Ding 1 Bisey
Bisey, Zaïre; 18.8°E, 4.6°S
Mukiramfi (Mukash-Kalel)

A search for "Bisey" (or alternative spellings) did not produce any results in geonames.org. We found a place called Biseye but the coordinate values (-4.08, 19.03) did not match those in (3). At this point, we inserted the coordinate values in (3) in the latitude and longitude tabs on the website gps-coordinates.net. This website is mainly used to convert coordinates from D(egree) M(inute) S(econd) to D(ecimal) D(egree) formats and vice-versa, but it also provides a coarse-grained map to visualize the location of the dot formed by a set of coordinates in relation to major cities within a given country. For example, -4.6, 18.8 in (3) is a place located southwest of the city of Nkara in the current Kwilu province of the DRC (see Map 1). We then searched in geonames.org, which offers extremely detailed, fine-grained maps with lots of place names, a place southwest of the city of Nkara in the DRC that would have coordinates closely matching -4.6, 18.8 in (3). Our best match was a place called Kwilumpia located at -4.61, 18.86. As a result, we renamed #92 Kwilumpia because this place can actually be located on a digital map and we could verify its coordinates.

16. In fact, variety names for North Teke entries (#16-20) in the original source are inconsistent: only #18 is named in the original source after a department (Bouenza). Entries #16, 17, 19 and 20 are named in the original source after a specific town/city. We tried to uniformize these entries by naming all of them after the Republic of the Congo department where the data originate.

17. Confusingly, the format in which geographical coordinates are presented in Bastin *et al.* (1999) is a mixture of D(egree) M(inute) S(econds) format and D(ecimal) D(egree) format. If the set of coordinates in (3) were expressed in DD format, they would read: -4.6, 18.8. If the set of coordinates in (3) were expressed in DMS format, they would read: E 18° 48' S 4° 35'. Unlike Bastin *et al.* (1999), we use the DD format 18.8, -4.6 in this study.

We applied this same procedure to cases where Bastin *et al.* (1999) do not provide the name of a geographical location along with a set of coordinates. Consider #91 West Ding (Kamtsha, Itere) which in the original source (Bastin *et al.* 1999: 13) appears as in (4).

(4) B86 Ding 2
Zaïre; 19.3°E, 4.3°S
Mundeke Otom'si Ebok, 1990

In this case too, we tried to locate a place geographically as close as possible to the coordinates -4.3, 19.3 in (4). The town of Itere in the current Kwilu province of the DRC is located at -4.28, 19.33. Because this is the closest place we could find given the coordinates in (4), we added Itere to the variety name in #91.

We proceeded this way also in some cases where Bastin *et al.* (1999) indicate more than one place name for a doculect, including a main town/city. Consider #89 named in the original source Ding 3 Sedzo and renamed by us West Ding (Mbentsie, Mpume). Information on this variety in the original source (Bastin *et al.* 1999: 13) appears as in (5).

(5) Ding 3 Sedzo
Sedzo, Ipaamvu, Zaïre; 19.2°E, 4.1°S
N. Burssens 1990

Sedzo is a well-known geographical location in the Kwilu region of the DRC (see Map 1). Its coordinates are -4.04, 19.17. A search for "Impaamvu" on geonames.org did not yield any results. We found a place called Mpume located at -4.12, 19.25 and we renamed the doculect in #89 after this town, because its latitude value (-4.12) is closer to the latitude value in the original source (-4.1) than the latitude value of Sedzo (-4.04).[18]

Finally, we discuss entries shaded in gray across the board in Table 2 (i.e. entries #62, 69, 90, 94 and 99). Recall that these are the entries for which we found data in the Excel file we obtained from Yvonne Bastin but which are not listed in the publication of Bastin *et al.* (1999). For these, we had no way of determining where the data were collected or originated. To solve this problem, we randomly selected a place within the area where we know a given variety to be spoken and assigned a set of coordinates to it.[19]

18. Another convenient reason for choosing Mpume instead of Sedzo is that many other varieties in this study are spoken in and around Sedzo and this would result in too many dots on this location on a map, hindering the visualization of doculect locations.

19. In a couple of cases, we relied on potential hints found in the Excel file we obtained from Yvonne Bastin. For example, the data for #90 is reported to have been collected by J. Daeleman in 1978 with a native speaker whose name is reported as "Ebaal". We thought that Ebaal might be an abbreviation of Ebalansthim Masuwan who wrote a grammatical sketch of Ding under the direction of Jan Daeleman. Ebalantshim Masuwan (1980: 2) describes a variety of Ding (B86) spoken in the collectivity of Bulwem, which is part of the collectivities of Sedzo and Matar, both located in the northern part of an area known as Kamtsha, Kamtshia or Kaantsa, among others, depending on the source. As a result of

A.2 Specifying available regiolectal information for language varieties

Whenever possible, we consulted additional available sources to include dialectal information doculects. Consider #88-93 in Table 2, where we added East and West to the Ding (B86) data following dialectal information found in Boone (1973). According to Boone (1973: 20), speakers of Ding come in two groups: those located west of the Mpio-Mpio River (-4.38, 19.6) in the current Kwilu province of the DRC and those located east of this river. We established coordinates for each of the places where varieties #88-93 were collected or originated so that we could locate each place east or west with respect to the longitude value of the Mpio-Mpio River. For instance, because Bambudi is located east of the Mpio-Mpio River at -4.30, 19.92, we named entry #88 East Ding (MBENTSIE, Bambudi). Similarly, because Kwilumpia is located west of the Mpio-Mpio River at -4.61, 18.86, we named #92 West Ding (KAMTSHA, Kwilumpia). The terms MBENTSIE and KAMTSHA in small caps before place names in entries #88 through #93 indicate how the eastern and western regiolects of Ding (B86) are commonly known in the literature (see for instance Mertens 1939; Ebalantshim Masuwan 1980). The majority of sources (van Bulck 1948: 177; Mwan Mesongolo 1984; Koni Muluwa & Bostoen 2015: 20) label the eastern group Mbentsie, but some call it Mukene (Mertens 1939), Munken (Mula 1977) or Munkeen (Mufanga-Dzmar 1977).[20] We now discuss the source of each dialectal specification in Table 2.

Entries #58-65: both van Bulck (1948: 485, 1954: 260) and Swartenbroeckx (1948: 6) identify an eastern vs. western divide of Yans (B85) varieties. We relied on geographical information provided in these sources and on the impressionistic line drawn on a map in Swartenbroeckx (1948: 6) to divide Yans varieties into eastern and western. The only other dialectal study of Yans varieties known to us is Mayanga (1985), who establishes over 40 distinct Yans dialects spread over 11 "collectivités" and names them after place names rather than dividing them into bigger groups based on cardinal points or other criteria.

Entries #66-68: this dialectal distinction is based on de Beaucorps (1941), van Bulck (1948) and Boone (1973). The term LUNIUNGU for East Nsong and GOBARI for West Nsong are originally attributed to Nsong speakers in de Beaucorps (1941: 5 and ff.). The Nsong are found in two distinct blocks in the current Kwilu province of the DRC, separated by Mbala speakers, among others (see also Vansina 1966: 131). De Beaucorps (1941) calls the west block *tsong de la Gobari*. This group extends mainly on the left side of the Gobari River in the territory of Masi Manimba (see Map 1) and on the right side of the Gobari River, in the territory of Kikwit (sector of Kwilu-Gobari). De Beaucorps (1941) calls the east block *tsong de la Luniungu*. This groups spreads over the territory of Kikwit (see Map 1), in the sectors of Basongo-Bambala, Kwilu and Nko.

---

this inferential reasoning, we labeled #90 West Ding (KAMTSHA, Sedzo). Similarly, #62 was labelled in the Excel file "Yanz Kumakuma". We did a search on geonames.org and found a place named Longo Kuma-Kuma in the area where Yans (B85) is spoken. We then renamed the entry East Yans (Longo Kumakuma).

20. The Ding spoken in the Kamtsha area is also called Ding Lesye (Mwan Mesongolo 1984). Depending on the source, Kamtsha can be spelled Kamtshia, Kaantsa or Kaansa among others.

Entries #76-78: these dialectal distinctions are based on Khang Levy (1979), which is a description of the Lwel spoken in Sedzo and Mateko in the DRC. In his description, Khang Levy (1979: X) presents a map distinguishing two dialectal groups of Lwel speakers, which he labels *lwel occidentaux* and *lwel orientaux*. These two groups are geographically separated by two groups of Ngwi speakers, which he labels *ngwi occidentaux* and *ngwi orientaux*.

Entry #96 West Mbuun (Imbongo): this dialectal distinction is based on Dibata Mimpiya (1977) and Koni Muluwa (2010). We refrained from assigning further dialectal specifications to Mbuun varieties in #94-95 and #97-99 in Table 2 to avoid confusion. This is because non-west dialectal subdivisions for Mbuun (B87) vary considerably depending on the author (see also Koni Muluwa & Bostoen 2015: 20). For instance, the Mbuun spoken in Idiofa is called the Central variety of Mbuun by Mundeke (1977, 1979) and the North variety of Mbuun by Dibata Mimpiya (1977).

A.3 Mismatches between place names and coordinates

When checking geographical data from Bastin *et al.* (1999) for B50-80 varieties (cf. Section 1.3), we encountered several problems and mismatches between place names and coordinates. We discuss these below.

Entry #37 Fumu (DRC): In Bastin *et al.* (1999) this variety is said to originate in Kwamouth, DRC at -3.2, 15.3. The coordinates for Kwamouth in the DRC are -3.18, 16.2. The coordinates -3.2, 15.3 point to a place in the Republic of the Congo, north of Brazzaville, in the Lesio-Louna Wildlife Reserve. We replaced original coordinates with coordinates for Kwamouth in the DRC.

Entry #38 Wuumu (Kimwenza): This variety is reported in Bastin *et al.* (1999: 12) as originating in Isaangampyu ('Kiùmwaansa'), DRC, at -4.5, 15. These coordinate values point to a place in the Republic of the Congo, southwest of Brazzaville. A search for "Isaangampyu" on geonames.org did not produce any results. However, we found a place called Kimwenza, located at -4.45, 15.28, southeast of Kinshasa, in an area where Wuumu is reported to be spoken in several sources (Guthrie 1953; Vansina 1966: 131; Boone 1973: 230). We relied on the possibility that Kiùmwaansa in Bastin *et al.* (1999) is an alternative spelling of Kimwenza.

Entry #46 Tiene (Dya, Mansele): This variety is reported in Bastin *et al.* (1999: 13) as being spoken in Mansele, DRC at -2.7, 17.7. We found a place called Mansele northeast of the town of Bolobo (in the current DRC province of Mai-Ndombe, see Map 1) in the area where the Dya dialect of Tiene is spoken (cf. Ellington 1977: xi), with coordinate values -2.08, 16.45. The coordinates -2.7, 17.7 point to a place close to Nioki, northeast of Mushie (province of Mai-Ndombe), close to Lake Mai-Ndombe. Since no source reports Tiene to be spoken in this area, we decided to stick to Mansele and changed the coordinate values accordingly.

Entry #53 North Boma (Mpukumbu): In Bastin *et al.* (1999: 13) this doculect is reported to originate in the DRC in the towns of Boku and Kondzulu. Coordinate values are reported as -3.5, 16.15. While we could not find coordinates for Kondzulu,

Boku is a town in the DRC located at -3.89, 16.63. This place is located in the vast region south of the Kwa River, between the Kwango and the Congo Rivers, northeast of Kinshasa (see Map 1). The coordinates in Bastin *et al.* (1999: 13), i.e. -3.5, 16.15, also point to a place located in this vast region, but much more westwards with respect to Boku, just on the right side of the Congo River. By looking at the wordlist for entry #53, we are pretty confident that #53 is a variety of North Boma (B82) and not of other Boma varieties (cf. Section 3.2) spoken in the vast region south of the Kwa River and northeast of Kinshasa. However, to our knowledge, no source reports North Boma (B82) to be spoken in the area to which Boku and the coordinate values -3.5, 16.15 belong. To solve this mismatch, we randomly selected a set of coordinates around Mushie in the current DRC province of Mai-Ndombe, because this is the main area where North Boma (B82) is spoken (cf. Guthrie 1953; Tonnoir 1970; Sulzmann 1983; Stappers 1986, amongst others).

Entry #55 Mfinu (Yuo): In Bastin *et al.* (1999: 13), this variety is reported to originate in Yuo, DRC, at -5.3, 14.7. The coordinates for a place called Yuo southeast of Kinshasa in an area where Mfinu (B83) is reported to be spoken by several sources (Guthrie 1953; Vansina 1966; Boone 1973) are -4.71, 16.06. Coordinate values -5.3, 14.7 point to a place southwest of Mbanza Ngungu in the current DRC province of Kongo-Central. This place appears to be way too southwest to host Mfinu speakers according to our sources. Consequently, we substituted the coordinates for this entry in Bastin *et al.* (1999: 13) with the ones found on geonames.org for the town of Yuo in the region of Kinshasa.

Table 2. B50-80 varieties included in this study[21]

#	Variety	Variety (or. source)	C o d e G/M	Code	Source	Coordinates (or. source)		Coordinates		Reference place
						N/S	E	N/S	E	
1	Wanzi (Kessipougou)	Duma 2 d. wanzi	B501	B501	B [B51/2]	-0.9	12.8	-0.9	12.78	Kessipougou (G)
2	Vili (Sindara)[22]	Vili	B503	B503	DDL/ ALGAB			-1.03	10.65	Sindara (G)
								-0.7	10.24	Lambaréné (G)
3	Duma (Lastoursville)	Duma 1	B51	B51	B [B51/1]	-0.8	12.6	-0.8	12.6	Lastourville (G)
4	Nzebi (Congo)	Nzebi 2 Congo	B52	B52X	B [B52/2]	-2.6	12.1	-2.67	12.08	Divenié (C)
5	Nzebi (Gabon)	Nzebi 1 Gabon	B52	B52Y	B [B52/1]	-1.9	11.9	-1.9	11.9	Mbigou (G)
6	Tsaangi (Madouma)	Tsangi	B53	B53	B [B53]	-2.9	13.7	-2.98	12.74	Madouma (C)

21. Abbreviations in Table 2 are as follows: ALGAB=Atlas linguistique du Gabon; B=Bastin *et al.* (1999); C=Republic of the Congo; DDL=Laboratoire Dynamique du Language (Université Lumière Lyon 2); DRC=Democratic Republic of the Congo, E=east; G=Gabon (only when in between parentheses); G=Guthrie (1971); KM&B=Koni Muluwa and Bostoen (2015); M=Maho (2009); N/A= does not apply; NN=Nsuka Nkutsi (1990); N/S=north/south; or. =original; St=Stappers (1986); Sw=Swartenbroeckx (1948).

22. We did not have access to any geographical information concerning this variety. We established coordinates based on Idiata *et al.* (2013). However, an anonymous reviewer observes that he has never found Vili (B503) speakers located so westwards in Gabon.

7	Kaning'i (Masuku)	Kaningi	B602	B602	B [B64]	-1.75	13.6	-1.75	13.6	Franceville (G)
8	Mbete (Ambinda)[23]	Mbere	B61	B61	B [B61]	-0.4	14	-0.4	14	Ambinda (G)
9	Mbaama (Gabon)	Mbamba 1 Gabon	B62	B62X	B [B62/1]	-1.2	13.9	-1.1	13.87	Ossélé (G)
10	Mbaama (Congo)[24]	Mbamba 3 Gabon or Mbamba 2d. mpini	B62	B62Y	B [B62/2] or [B62/3]	-0.3	14.4	-0.33	14.48	Léboka (C)
11	Nduumo (Yéyé)	Ndumo 2 d. kuya	B63	B63X	B [B63/2]	-1.5	13.4	-1.49	13.4	Yéyé (G)
12	Nduumo (Franceville)	Ndumo 1	B63	B63Y	B [B63/1]	-1.6	13.5	-1.6	13.5	around Franceville (G)

23. An anonymous reviewer observes that he has never found speakers of Mbete (B61) located in eastern Gabon. However, the map in Lane (1989: 11) clearly shows that Mbete (B61) and Mbaama (B62) varieties are spoken in eastern Gabon and in western and southwestern parts of the Republic of the Congo, stretching over the border between the two countries.

24.Entry #10 appears in the Excel file of Bastin *et al.* (1999) as "mbaamba" and the transcriber as "Bouka 1989?". However, in the published version of Bastin *et al.* (1999) there are two datasets that could correspond to this entry. These are:

a) B62 Mbamba 2 d. mpini
Congo; 14.5°E, 1.2°S
L. Y. Bouka, 1989?

b) B62 Mbamba 3 Gabon
(d. de Kelle et Ewo); Congo; 14.4°E, 0.3°S
L. Y. Bouka, 1989?

"Gabon" in dataset b) is very likely a typo. Because both sets of coordinates in a) and b) point to locations in the Republic of the Congo, we named entry #10 "Mbaama (Congo)".

13	Bibaana (Dumu)	Bibaana		B70x	NN			-3.46	16.73	Dumu (DRC)
14	South Teke (Ngi)	Ngi (Teke-sud)		B70y	NN			-4.1	15.89	Ngi (DRC)
15	Bwala (Bankana)	N/A	N/A	B70z	2018 Fieldwork by Flore Bollaert (UGent)	N/A	N/A	-4.45	16.19	Bankana (DRC)
16	North Teke (Plateaux)	Tee 2 Abala	B71a	B71a	B [B71/2]	-1.3	15.4	-1.28	15.44	Eko (C)
17	North Teke 2 (Cuvette)	Tee 4 Omvula	B71	B71W	B [B71/4]	-1.7	14.5	-1.65	14.53	Assiéné (C)
18	North Teke (Bouenza)	Tee 3 Bouenza	B71	B71X	B [B71/3]	-3.7	14.1	-3.73	14.14	Kingoué (C)
19	North Teke 1 (Cuvette)	Tee 1 Boundji	B71	B71Y	B [B71/1]	-1.1	15.3	-1.1	15.35	Boundji (C)
20	North Teke (Pool)	Tee 5 Mbe	B71	B71Z	B [B71/5]	-3.7	14.5	-3.72	14.52	Kindamba (C)
21	Ngungwel (Gamboma)	Ngungwel	B72a	B72a	B [B72]	-1.9	15.8	-1.9	15.8	Gamboma (C)
22	Tsaayi (Bambama)	Teke tsaayi/ Teke-	B73a	B73a	B [B73/2]	-2.6	13.5	-2.55	13.54	Bambama (C)

		W 2 Bambana[25]								
23	Laali (Bouenza)	Teke-W 5 d. lali Bouenza	B73b	B73bX	B [B73/5]	-3.7	13.9	-3.7	13.9	Tsiangi (C)
24	Laali (Lekoumou)	Teke-W 6 d. lali Lekoumou	B73b	B73bY	B [B73/6]	-3.7	13.3	-3.7	13.3	Sibiti (C)
25	Yaa (Bihoua)	Teke-W 4 d. iyaa	B73c	B73c	B [B73/4]	-3.8	13.3	-3.81	13.32	Bihoua (C)
26	West Teke (Komono)	Teke-W 1 Komono	B73	B73X	B [B73/1]	-3.3	13.2	-3.27	13.22	Komono (C)
27	West Teke (Kissiélé)	Teke-W 3 Kissiele	B73	B73Y	B [B73/3]	-2.8	12.6	-2.76	12.64	Kissiélé (C)
28	Eboo-Nzikou (Ngo)	Boma	B74a, B74b	B74[26]	B [B74]	-2.5	15.6	-2.5	15.6	Ngo (C)
29	Tio Bali (Ibali)	Teke	B76	B75W[27]	KM&B			-3.78	17.22	Menkwo (DRC)
30	Tio Bali 2 (Tua)	Teke-E 2 Tua	B76[27]	B75X	B [B76/2]	-3.65	16.6	-3.63	16.61	Tua (DRC)

25. The place name Bambana in the original source is either a misspelling or an alternative name for "Bambama". Whatever the case might be, the coordinates available in Bastin *et al.* (1999) for this variety match the coordinates of a place called Bambama in the DRC.

26. Bastin *et al.* (1999) distinguishes between two dialectal variants of B74 Central Teke, B74a Njyunjyu/Ndzindziu and B74b Boo, Boma. However, Raharimanantsoa (2012a) shows that B74a and B74b are phonologically essentially the same variety. This is why we use B74 for this entry instead of B74a or B74b.

27. For a discussion of why we assigned the code B75 to varieties previously numbered B76 see Section 3.4.

31	Tio Bali 1 (Tua)	Teke-E 3	B76	B75Y	B [B76/3]	-3.6	16.65	-3.6	16.65	Tua (DRC)
32	Tio Bali (Fatundu)	Teke-E 1	B76	B75Z	B [B76/1]	-4.1	17.2	-4.11	17.22	Fatundu (DRC)
33	Mosieno (Bolobo)	siene/Teke-E 4 Bolobo	B76a	B76a	B [B76/4]	-2.2	16.2	-2.15	16.23	Bolobo (DRC)
34	Kukwa (Plateau)	Kukwa 1	B77a	B77aX	B [B77/1]	-2.6	14.3	-2.6	14.3	Plateau Koukouya (C)
35	Kukwa (Lekana)	Kukwa 2	B77a	B77aY	B [B77/2]	-2.3	14.5	-2.3	14.57	Lekana district (C)
36	Fumu (Congo)	Fumu-Wumbu 1d. fumu Congo	B77b	B77bX	B [B78/1]	-4.2	15.3	-4.2	15.3	Brazzaville (C)
37	Fumu (DRC)	Fumu-Wumbu 2 d. fumu Zaïre	B77b	B77bY	B [B78/2]	-3.2	15.3	-3.18	16.2	Kwamouth (DRC)

38	Wuumu (Kimwenza)	ruumbu Kimwaansa/ Fumu-Wumbu 4 d. wumbu Isaangampyu	B78	B78W	B [B78/4]	-4.5	15	-4.45	15.28	Kimwenza (DRC)
39	Wuumu (Kasangulu)	Fumu-Wumbu 6 d. wumbu Kasangulu	B78	B78X	B [B78/6]	-4.6	15.15	-4.58	15.16	Kasangulu (DRC)
40	Wuumu (Mobenga)	Fumu-Wumbu 3 d. wumu Mobenga	B78	B78Y	B [B78/3]	-4.3	17.05	-4.34	17.02	Mobenga (DRC)
41	Wuumu (Mambulu)	Fumu-Wumbu 5 d. wumbu-E	B78	B78Z	B [B78/5]	-4.4	16.9	-4.39	16.94	Mambulu (DRC)
42	Boma Nkuu (Monkana)	Nkuu		B80x	NN			-3.42	17.26	Monkana (DRC)
43	South Boma [Nkuu] (Boku)	Boma		B80y	NN			-3.89	16.63	Boku (DRC)

44	Boma Yumu (Pentane/ Mondai)	Boma	B82	B80zX	KM&B			-3.27	17.45	Kimbari (DRC)
								-3.35	17.49	Ito (Yenge Yenge) (DRC)
								-3.32	17.57	Mondai (DRC)
								-3.44	17.6	Mushie-Pentane (DRC)
45	Boma Yumu (Saio)	Boma 2 Saio	B82	B80zY	B [B82/2]	-3.8	18.2	-3.79	18.18	Saio (DRC)
46	Tiene (DYA, Mansele)	Tiene 2 d. Dya	B81	B81X	B [B81/2]	-2.7	17.7	-2.08	16.45	Mansele (DRC)
47	Tiene (Wamba)[28]	Tiene 3 Wamba	B81	B81Y	B [B81/3]	-4.3	15.6	-4.36	15.8	Wamba (DRC)
48	Tiene (Bosiki)	Tiene 1 Botala	B81	B81Z	B [B81/1]	-2.8	17	-2.75	17.03	Bosiki (DRC)
49	Mpe (Ile)	Kempee	B821	B821	B [B80K]	-2.3	18	-2.25	17.95	Ile (DRC)

28. This doculect originates in an area where Tiene is not reported to be spoken. As we show in Section 4, this doculect appears to be heavily influenced by contact with other varieties.

50	Nunu (Letomo)	Nunu	B822	B822	B [B80Nu]	-3.1	16.9	-3.05	16.87	Letomo (DRC)
51	North Boma (Mushie)	Boma	B82	B82X	St			-3.01	16.92	Mushie (DRC)
52	North Boma (Mbali-Iboma)	Boma 1 Ekemwa	B82	B82Y	B [B82/1]	-2.4	17.25	-2.38	17.28	Mbali-Iboma (DRC)
53	North Boma (Mpukumbu)	Boma 3 Kondzulu	B82	B82Z	B [B82/3]	-3.5	16.15	-2.89	16.66	Mpukumbu (DRC)
54	Mfinu (Bita)	Mfinu 3 Mungata	B83	B83X	B [B83/3]	-4.2	15.8	-4.22	15.82	Bita (DRC)
55	Mfinu (Yuo)	Mfinu 1	B83	B83Y	B [B83/1]	-5.3	14.7	-4.71	16.06	Yuo (DRC)
56	Mpuono (Kindunu)	Mpuono 1	B84	B84X	B [B84/1]	-4.4	15.7	-4.4	15.67	Kindunu (DRC)
57	Mpuono (Fadiaka)	Mpuono 2 Fadiaka	B84	B84Y	B [B84/2]	-3.95	16.9	-3.96	16.9	Fadiaka (DRC)
58	West Yans (Mukonkie)	Yans Yey	B85a	B85a	Sw			-3.48	17.29	Mukonkie (DRC)
59	East Yans (Nkara)	Yans	B85b	B85bT	KM&B			-4.53	18.89	Nkara (DRC)
60	East Yans (Kiwanda)	Yanz 5	B85b	B85bU	B [B85/5]	-4.7	18.7	-4.73	18.72	Kiwanda (DRC)

61	East Yans (Mansthiene)[29]	Yanz 3	B85b	B85bV	B [B85/3]	-4.5	18.5	-3.81	17.71	Mantshiene (DRC)
62	East Yans (Longo Kumakuma)	Yanz Kumakuma [not in publ. ver.]	B85b	B85bW	B [Fungula /Vansina]			4.61	18.93	Tshama near Longo Kuma-Kuma (DRC)
63	East Yans (Bagata)	Yanz 6 d. yey Bagata	B85b	B85bX	B [B85/6]	-3.7	17.95	-3.72	17.95	Bagata (DRC)
64	East Yans (Kibongo)	Yanz 1 Kibongo	B85b	B85bY	B [B85/1]	-4.4	18.7	-4.42	18.72	Kibongo (DRC)
65	East Yans (Kimbanda)	Yanz 2	B85b	B85bZ	B [B85/2]	-4.1	18.1	-4.08	17.98	Kimbanda (DRC)
66	West Nsong (GOBARI, Kikondji)	Yanz 8 d. tsong Masi	B85d	B85dX	B [B85/8]	-4.7	18.4	-4.73	18.4	Kikondji (DRC)
67	East Nsong (LUNIUNGU, Mwilambongo)	Yanz 9 d. tsong Mwilabongo	B85d	B85dY	B [B85/9]	-4.9	19.85	-4.95	19.85	Mwilambongo (DRC)

29. Bastin *et al.* (1999) do not specify a location for these data. They indicate that the data is from the DRC and they provide coordinates 18.5, -4.5 preceded by an asterisk. Bastin *et al.* (1999) explain that asterisks are used when they had no exact information (i.e. a place name) where the wordlist was recorded. In these cases, Bastin *et al.* (1999: 9) gave an arbitrary set of coordinates within the area where the language of the wordlist in question is spoken. Because they indicate that Rottland collected the data for entry #61, we consulted Rottland (1977) and selected one of the place where he claims to have collected the Yans data.

68	East Nsong (LUNIUNGU, Kipuka)	Nsong	B85d	B85dZ	KM&B			-5.09	18.81	Kipuka (DRC)
69	Mpur (Due I)	Mput [not in publ. ver.]	B85e	B85eW	B [Ideyi/ Angenot]			-4.33	18.84	Due I (DRC)
70	Mpur (Kwebe)	Mpur	B85e	B85eX	KM&B			-4.32	18.88	Kwebe (DRC)
71	Mpur (Due)	Yanz 10 d. mput Dwe	B85e	B85eY	B [B85/10]	-4.3	18.8	-4.29	18.82	Due (DRC)
72	Mpur (Kolonzadi)	Yanz 12 d. mput Kolonzadi	B85e	B85eZ	B [B85/12]	-4.3	18.9	-4.3	18.9	Kolonzadi (DRC)
73	Nsambaan (Kwilumpia/ Longo Kuma-Kuma)	Nsambaan	B85F	B85FX	KM&B			-4.61	18.86	Kwilumpia (DRC)
								-4.54	18.91	Longo Kuma-Kuma (DRC)

74	~~Nsambaan (Kwenge)~~ Samba[30]	~~Yanz 7 d. tsambaan Kimatu~~	~~B85F~~ L12a	~~B85FY~~ L12a	B [B85/7]	-4.9	18.6	-4.85	18.64	Kwenge (DRC)
75	Nsambaan (Nkara)	Yanz 13 d. tsambaan Nkara	B85F	B85FZ	B [B85/13]	-4.5	18.9	-4.53	18.89	Nkara (DRC)
76	East Ngwi (Mangai)	Ngwi	B861	B861X	KM&B			-4.05	19.57	Mangai (DRC)
77	West Ngwi (Mateko)	Ngwi 5 Mateko	B861	B861Y	B [B88/5]	-4.15	19.1	-4.15	19.08	Mateko (DRC)
78	East Lwel (Sedzo)	Lwel	B862	B862X	KM&B			-4.04	19.17	Sedzo (DRC)
79	Mpiin (Bamba)	Mpiin 1 Madzing	B863	B863X	B [B80Mp1]	-5.8	18.4	-5.76	18.37	Bamba (DRC)

30. This wordlist, collected in 1975 from a consultant called Kingunza by the Flemish Jesuit missionary and linguist Jan Daeleman (1922-2014), is available on the website of the Royal Museum for Central Africa (cf. https://www.africamuseum.be/sites/default/files/media/docs/research/human-sciences/culture-society/lexico_bantu/B/B851_kitsamb.pdf), where it is not coded as B85F, but as B851. Both codes are wrong, as was its inclusion in the lexicostatical study of Bastin *et al.* (1999) as a variety of Tsambaan or Nsambaan (i.e. their doculect B85/7). In the online wordlist, this doculect is labelled "Kitsamb". This doculect does not represent Nsambaan (B85F), but Samba (L12a). After being updated by Joseph Koni Muluwa, it was incorporated as Samba L12a in the phylogenetic study of de Schryver *et al.* (2015). We realized this after the peer review process. We decided to keep this doculect as B85FY and to cross it out in Table 2 to prevent future researchers from making the same mistake again.

80	Mpiin (Kipuka)	Mpiin	B863	B863Y	KM&B			-5.09	18.81	Kipuka (DRC)
81	Mpiin (Kikwit)	Mpiin 2 Biyung	B863	B863Z	B [B80Mp2]	-5	18.7	-5.04	18.71	Kikwit (DRC)
82	Ngong (Kwenge)	Ngong	B864	B864X	KM&B			-4.85	18.64	Kwenge (DRC)
83	Ngong (Makenge)	Ngongo-W 2 Kinkaam	B864	B864Y	B [B80Ng2]	-5.8	18.1	-5.81	18.13	Makenge (DRC)
84	Ngong (Kiluengo)	Ngongo-W 1 Kikasan	B864	B864Z	B [B80Ng1]	-5.7	18.2	-5.7	18.21	Kiluengo (DRC)
85	Nzadi (Indolo)	Nzadi	B865	B865X	KM&B			-4.12	19.89	Indolo (DRC)
86	Nzadi (Panu)	Nzadi 2 Panu	B865	B865Y	B [B80Nz2]	-3.8	19.1	-3.79	19.11	Panu (DRC)
87	Nzadi (Makanga)	Nzadi 1 Kutu	B865	B865Z	B [B80Nz1]	-3.7	19	-3.76	18.94	Makanga (DRC)
88	East Ding (Mbentsie, Bambudi)	Ding	B86	B86U	KM&B			-4.3	19.92	Bambudi (DRC)
89	West Ding (Mbentsie, Mpume)	Ding 3 Sedzo	B86	B86V	B [86/3]	-4.1	19.2	-4.12	19.25	Mpume (DRC)

90	West Ding (KAMTSHA, Sedzo)	Dzing (Kaantsa) [not in publ. ver.]	B86	B86W	B [Ebaal/J. Daeleman 1978]			-4.04	19.17	Sedzo (DRC)
91	West Ding (KAMTSHA, Itere)	Ding 2	B86	B86X	B [86/2]	-4.3	19.3	-4.28	19.33	Itere (DRC)
92	West Ding (KAMTSHA, Kwilumpia)	Ding 1 Bisey	B86	B86Y	B [86/1]	-4.6	18.8	-4.61	18.86	Kwilumpia (DRC)
93	West Ding (KAMTSHA, Oveke-Fiote)	Ding 4	B86	B86Z	B [86/4]	-4.2	19.4	-4.19	19.37	Oveke-Fiote (DRC)
94	Mbuun (Gungu)	Mbuun [not in publ. ver.]	B84	B87U	B [Kapumbu/ Mukash-Kalel]			-5.73	19.31	Gungu (DRC)
95	Mbuun (Milundu)	Mbun 3 Milundu	B84	B87V	B [B87/3]	-4.3	19	-4.25	19.01	Milundu (DRC)
96	West Mbuun (Imbongo)	Mbuun	B84	B87W	KM&B			-4.92	19.04	Imbongo (DRC)
97	Mbuun (Mayungu)	Mbun 2	B84	B87X	B [B87/2]	-4.5	19.5	-4.5	19.51	Mayungu/a (DRC)
98	Mbuun Mwilambongo	Mbun 1 Mwilambong	B84	B87Y	B [B87/1]	-4.9	19.8	-4.93	19.8	Mwilambongo (DRC)
99	Mbuun (Mabangi)	Mbuun Labaempi	B84	B87Z	B [Vansina]			-5.76	19.3	Mabangi (DRC)

Table 3. Evolutions in naming systems for current Guthrie B80 group[31]

Guthrie (1948: 33) zone B Group 40	**Guthrie (1953: 81) Tende-Yanzi Group**	**Guthrie (1970: 12) Tende-Yanzi Group [Congo-Kinshasa]**	**Maho (2009: 24) B80: Tiene-Yanzi Group**
41 MFINU	**B81** *TỊINI*, ki- [Tende]	**B81** Tịẹnẹ (Tende)	**B81** Tiene^{tii}, Tende
42 BOMA	**B82** *BOMA*, i-	**B82** Boma (Buma)	**B82** Boma^{boh}, Buma
			B821 Mpe, Kempee
			B822 Nunu
43 TIENE	**B83** *MFINU*, e- [Mfunika, Mfununga]	**B83** Mfinu (Funika, Mfununga)	**B83** Mfinu^{zmf}, Funika, Mfununga
44 SAKATA	**B84** *DJỊA* and *SAKATA*	**B84** Mpuono	**B84** (=B87) Mpuono^{zmp}, Mpuun, (Mbuun?), (Mbunda?)
	B84a *DJỊA*, ki-	**B84a** Mpuono proper	
	B84b *SAKATA*, ki-	**B84b** Mpụun	

31. In Table 3, we report all conventions (uppercase, lowercase, special characters, etc.) as found in the original source. We note that one and the same item listed in *The Key List of Bantu Languages* (Guthrie 1970: 11 and ff.) often appears slightly different in *An Inventory of Bantu Languages* in Guthrie (1971: 28 and ff.). For instance, Mpuono appears as Mpuon and Mbiem is followed by (W. Yanzi) in Guthrie (1971: 38). Guthrie (1970: 17) places an alphanumeric code in parentheses (cf. (B87) Mbuun in the third column) to indicate that there is no data cited from that language in his *Catalogue of Common Bantu with Commentary* (see also Table 4). Shaded cells in Table 3 indicate lack of a variety in a referential classification.

45 YANZI	**B85** *YANS*, i- [Yanzi]	**B85** Yans (Yanzi) **B85a** Mbiem **B85b** E. Yans **B85c** Yeei **B85d** Ntsuo **B85e** Mpur	**B85** Yans[yns], Yanzi **B85a** Mbiem, West Yansi **B85b** East Yans **B85c** Yeei **B85d** Nsong, Tsong[soo], Itsong, Ntsuo, "Songo" **B85e** Mpur, Mput
			B85F Tsambaan
46 Ŋgoli	**B86** *ŊGUL*, į- [Ngoli]		(see B861 below)
47 Diŋa	**B87** *DĬ*, i- and *DZĮŊ*, į-	**B86** Dį (Dinga, Dziŋ)	**B86** Di[diz], Dinga, Dzing **B861** Ngul[nlo], incl. Ngwi **B862** Lwel, Kelwer **B863** Mpiin, Pindi **B864** West Ngongo **B865** Nzadi
48 MBUNU	**B88** *MBUUN*, į- [Mbunda]	**(B87)** Mbuun (Mbunda)	~~**B87**~~ see Mpuono B84

Table 4. Some referential classifications of so-called Teke languages (a.k.a. current Guthrie's B70)[32]

Guthrie (1948: 33) Zone B Group 30	**Guthrie** (1953: 77) Teke Group	**Guthrie** (1970: 17) Teke Group	**Maho** (2009) B70: Teke Group	**Ethnologue** (Grimes 1992: 233, 432; 2000: 77, 103) Teke	**Ethnologue** (Lewis *et al.* 2016: 116 and ff., 138) Teke
32 Tege̱	**B71** N. TEG̱E and NJĮNĮIƆĮ **B71*a*** TEG̱E, ka- **B71*b*** NJĮNĮIƆĮ, ka-	**B71** N. Teke (C.B) **B.71a** Tɛgɛ-Kalį **B71b** Njįnįŋį	**B701** Tsitsege **B71** Teghe, North Teke **B71a** Keteghe, Tege-Kali **B71b** Kateghe, Njining'i, Nzikini	**Northern Teke** (Teghe, Iteghe, Tege) Kateghe (Nzikini) Keteghe	**Teke-Tege** (Iteghe, Keteghe, Northern Teke, Teghe, Teke Alima, Teke Kali) Kateghe (Nzikini) Keteghe
	B.72 MPŨ, ę- [Ngungulu]	**B72** N.E. Teke (C.B)	**B72(a)** Ngungwel, Ngungulu,	**North-eastern Teke** (Ngungwel, Ngungulu, Ngangoulou)	**Ngungwel** (Engungwel, Ngangalou, Ngangulu,

32. In Table 4, we kept all Guthrie's conventions (special characters, uppercase letters etc.) as found in the original source. In the column Guthrie (1970), C.B stands for Congo-Brazzaville (currently The Republic of the Congo) and C.K stands for Congo-Kinshasa (currently DRC). In the Ethnologue columns, names of main entries are bolded. Alternative names for main entries are in italics. Plain text is used for what the authors call dialects under each main entry. Unless otherwise specified all main entries in the Ethnologue columns refer to languages/groups found in The Republic of the Congo. Whenever possible, we assigned a random circled number to Teke varieties whose names matched between Guthrie (1953) and Guthrie (1970) so that the reader can more easily trace where a given variety ended up in Guthrie (1970), cf. ① next to B.73a in Guthrie (1953) and B.74b in Guthrie (1970). Although not included in Table 3, we refer the interested reader to Jacquot (1971) for an additional referential classification of Teke languages spoken in the Republic of the Congo. Some varieties of North Teke (B71) are also spoken in Gabon (Fontaney 1984; Idiata *et al.* 2013; Linton 2016). A shaded cell in Table 4 means that a language/group is lacking in a given classification.

		B.2a Ngungwel (Ngungulu) **B72b** Mpũmpũ	North-East Teke **B72b** Mpumpu	Mpu (Mpumpum)	Ngungulu, Norteastern Teke) Mpu (Mpumpu, Mpumpum)
33 Boma	**B73** BOÕ and KỤKẄA **B73*a*** BOÕ, ẹ- [Boma] ① **B73*b*** KỤKẄA②	**B73** W. Teke (C.B) **B.73a** Tsaayi **B73b** Laali ⑤ **B73c** Yaa (Yaka) ⑥ (B.73d) Kwe	**B73** West Teke **B73a** Tsaayi **B73b** Laali **B73c** Yaa, Yaka **B73d** Tyee, Tee, Kwe	**Western Teke** *(West Teke)* Tsaayi Laali Yaa Tyee	**Teke-Tsaayi** (Getsaayi, Tsaayi, Tsay-a/-e/-i, West Teke) **Teke-Laali** (Ilaali, Laali, West Teke) **Teke-Tyee** (Kwe, Tee, Tyee, West Teke) **Yaka** (Iyaka, West Teke, Yaa)
36 ƉEE	**B74** ƉEE, esi- ③	**B74** Central Teke (C.B.) **B4a** Ndzindziu **B74b** Boõ (Boma) ①	**B74** Central Teke **B74a** Njyunjyu, Ndzindziu **B74b** Boo, Boma	**Central Teke** *(Kiteke)* Tyoo Boo Njyunjyu **Central Teke** under "DRC" Kwe Njinju Wuo Boma	**Teke-Nzikou** (Central Teke, Ndzindziu, Njiunjiu, Njyunjyu) Teke-Eboo **[Teke-Nzikou]** also present under DRC

35 TIO	**B.75** TIO, i- [Teke]	**B75** Bali (Tio, Teke) (C.B, C.K)	**B75** Bali, incl. Teke, Tio	**Eastern Teke** *(Kiteke, Ibali)* **under DRC** Mosieno Ngee Bali (Ambali, Teo, Tio, Tyo)	**Teke-Ibali** (Bali, Eastern Teke, Ibali, Kiteke, Teke-Ibali) **under "DRC"** Mosieno Ngee (Esingee) Bali (Ambali, Teo, Tio, Tyo)
37 WUMU	**B.76** WUMU, i- [Wumbu, Mbunu] ④	B.76 East Teke (C.K.) **(B76a)** Mosieno B.76b Ŋee ③	**B76** East Teke **B76a** Mosieno **B76b** Ng'ee		
34 YAKA	**B.77** S.W. TEGE **B.77*a*** TEGE, i- **B.77*b*** LAALI, i- ⑤ **B.77*c*** Yaa, i- [Yaka] ⑥	B.77 S. Teke (C.B.) **B77a** Kukẅa ② B.77b Fumu ⑦	**B77a** Kukwa (South Teke) **B77b** Fu(u)mu (South Teke)	**Southern Teke** (Kukwa, Kukuya, Kikuwa, Chikuya, Koukouya)	**Teke-Kukuya** (Chikuya, Cikuya, Kikuwa, Kikuya, Koukouya, Kukwa, Kukuya, Southern Teke)
31 FUMU	**B.78** FUMU, i- ⑦	**B78** Wuumu (Wumbu) (C.K) ④	**B78** Wuumu, Wumbu	**South Central Teke** *(Kiteke)* Fuumu Wuumu	**Teke-Fuumu** (Fumu, Kiteke, South Central Teke, Teke du Pool) Fu(u)mu (Ifuumu, Mfumu) Wuumu (Iwuumu, Wumbu)

Author's addresses

Sara Pacchiarotti
UGent Centre for Bantu Studies (BantUGent)
Department of Languages and Cultures
Blandijnberg 2
9000 Ghent
Belgium
sara.pacchiarotti@UGent.be

Natalia Chousou-Polydouri
Department of Comparative Linguistics
University of Zürich
Plattenstrasse 54
CH-8032 Zürich
natalia.chousou-polydouri@uzh.ch

Koen Bostoen
UGent Centre for Bantu Studies (BantUGent)
Department of Languages and Cultures
Blandijnberg 2
9000 Ghent
Belgium
koen.bostoen@UGent.be

Résumé

Dans cet article, nous traitons de l'identification, de la géographie et de la phylogénie interne des langues bantu B50-80. Selon les études lexicostatistiques et phylogénétiques basées sur le lexique, ces langues appartiennent à la branche Côte-Ouest de la famille bantu, tout comme le groupe kikongo. Nous commençons par présenter une liste mise à jour des près de 100 variantes linguistiques des groupes B50-80 de Guthrie et des coordonnées géographiques correspondantes, mises à jour également. Nous dissipons la confusion qui entoure certains glossonymes et pointons du doigt les mauvaises interprétations qui ont conduit au regroupement de variétés en réalité distinctes, et à des erreurs de numérotation dans les classifications référentielles non génétiques (Guthrie 1971, Maho 2009, Hammarström 2019). Nous présentons ensuite les résultats d'une nouvelle étude phylogénétique de toutes les variantes B50-80 de notre échantillon, que nous comparerons plus tard à une classification interne réalisée sur base d'innovations phonologiques partagées. Les résultats que nous avons obtenus montrent que (i) les classifications internes des langues B50-80 doivent soit être révisées, soit être reconsidérées entièrement et (ii) la structure interne de la branche Côte-Ouest se dégageant de notre étude suggère que son berceau ne se trouve pas, comme on l'a cru précédemment, quelque part entre le plateau Batéké et la région du Bandundu en RDC, mais beaucoup plus à l'est, entre les rivières Kamtsha et Kasaï, toujours en RDC.